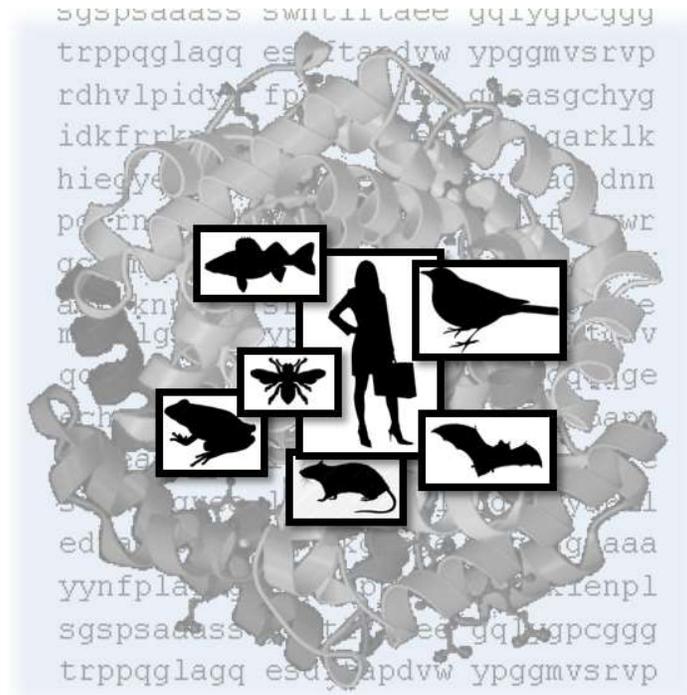


EPA/600/R-23/225

# Sequence Alignment to Predict Across Species Susceptibility

(SeqAPASS)

VERSION 7.0



## User Guide

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) User Guide

**Quick Notes:** Use [Chrome](#) for optimal performance and [PLEASE DO NOT submit more than 10 Level 1 or Level 4 queries at a time](#). Wait until jobs run to completion prior to submitting more.

### Table of Contents

Background	page 3
Accessing SeqAPASS	page 4-5
<i>Returning Users</i> (page 4)	
<i>First Time Users</i> (page 5)	
Messages from the SeqAPASS Development Team	page 5
SeqAPASS Home Tab	page 6
Request SeqAPASS Run Tab	page 6-12
<i>Identify a Protein Target</i> (page 7)	
<i>Query “By Species”</i> (page 8)	
<i>Query “By Accession”</i> (page 11)	
SeqAPASS Run Status	page 12-15
View SeqAPASS Reports	page 15-20
<i>View Report</i> (page 15)	
<i>Save Report(s)</i> (page 17)	
Level 1: Primary Amino Acid Sequence Alignment	page 21-26
<i>Primary Report Settings</i> (page 23)	
Susceptibility Cutoff Box for Level 1	page 27-29
<i>No Orthologs Detected</i> (page 28)	
ECOTOX Widget	page 30-31
Level 2: Functional Domain(s) Alignment	page 31-33
View Level 2 Data Page	page 34-38
<i>Primary Report Settings</i> (page 36)	
Susceptibility Cutoff Box for Level 2	page 39-42
<i>No Orthologs Detected</i> (page 41)	
Level 1 and Level 2 Data Visualization	page 42-51
<i>Level 1 and 2 Information Page</i> (page 43)	
<i>Level 1 and 2 BoxPlot Page – Controls</i> (page 45)	
Level 3: Individual Amino Acid Residue Alignment	page 51-63
View Level 3 Individual Amino Acid Query and Data Page	page 63-68
<i>Level 3 Data – Primary Report</i> (page 66)	
<i>Level 3 Data – Full Report</i> (page 67)	
Heat Map	page 69-73
Level 4: Protein Structural Analysis	page 73-93
<i>Level 4 FASTA Generation</i> (page 74)	
<i>Level 4 View FASTA</i> (page 75)	
<i>Level 4 View I-TASSER</i> (page 77)	
<i>Level 4 Page</i> (page 78)	
<i>Level 4 Input Other Protein Structures for TM-align: AlphaFold, RCSB PDB, and Other</i> (page 79)	
<i>Level 4 TM-align</i> (page 80)	

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Decision Summary Report	page 82-85
Download DS Report as PDF	page 86-88
Moving Between Level 1, Level 2, Level 3, and Level 4 Data Pages	page 88-89
Search, View, and Download Data Tables	page 89-90
Log out	page 90
Pop-up Messages	page 90-93
SeqAPASS Documentation	page 94-104

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Background

The SeqAPASS tool has been developed to predict across species relative intrinsic susceptibility to chemicals with known molecular targets (e.g., pharmaceuticals, pesticides) as well as evaluate conservation of molecular targets from high-throughput screening assays (i.e., U.S. Environmental Protection Agency ToxCast Program) and molecular initiating events (MIEs) and early key events in the adverse outcome pathway (AOP) framework, as a means to extrapolate such knowledge across species. The term “relative” is used because it is recognized that molecular target similarity is one consideration, though an important one, for making predictions of susceptibility to a chemical. Other important considerations for susceptibility that are not evaluated using the SeqAPASS methodology include how well a chemical is absorbed, distributed, metabolized, and eliminated, life stage, and other life history traits. Also, “relative” indicates that the determination of sequence similarity between proteins is based on comparison to a single protein sequence for a specific species. Additionally, we describe “intrinsic susceptibility” as the vulnerability (or lack thereof) of an organism to chemical perturbation due to its inherent biological composition.

Cross-species comparisons of proteins can be conducted through examination of sequence and structural information, depending on how well the protein has been characterized and what is known about a chemical-protein interaction. SeqAPASS allows the user to assess various levels of protein sequence detail across species including comparisons of primary amino acid sequence (including ortholog detection), functional domain(s), individual amino acid residue positions, and structural similarity. Each level requires a greater understanding of the protein and its interaction with a chemical of interest (or similar ligand).

Because human and veterinary drugs, as well as pesticides, are designed to act specifically on well characterized molecular targets, these chemical classes have proven useful for demonstrating the utility of the SeqAPASS tool and its application to various hazard assessment/research scenarios.

The pertinent information necessary to begin a SeqAPASS query includes: the identification of a single (or multiple) query species and a query protein, which would be the molecular target(s) of interest (e.g., receptor or enzyme).

The SeqAPASS algorithms mine, collect, and collate information from the National Center for Biotechnology Information (NCBI) protein database (<http://www.ncbi.nlm.nih.gov/protein/>), conserved domains database (<http://www.ncbi.nlm.nih.gov/cdd/>), taxonomy database (<http://www.ncbi.nlm.nih.gov/taxonomy/>), strategically utilizes the Stand-Alone Basic Local Alignment Search Tool for proteins (BLASTp) ([http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE\\_TYPE=BlastDocs&DOC\\_TYPE=Download](http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download)) and the Constraint-based Multiple Alignment Tool (COBALT) ([http://www.st-va.ncbi.nlm.nih.gov/tools/cobalt/re\\_cobalt.cgi](http://www.st-va.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi)).

Advanced users can request Level 4 access for protein structure generation and alignment to gather additional lines of evidence toward conservation. Tools used for structural evaluation include: Iterative Threading ASSEmblY Refinement (I-TASSER, <https://zhanggroup.org/I-TASSER/>) for creation of protein structural models, AlphaFold (<https://AlphaFold.ebi.ac.uk/>) and other sources using PDB formatted protein structures for structural alignment using TM-align (<https://zhanggroup.org/TM-align/>).

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: [LaLone.Carlie@epa.gov](mailto:LaLone.Carlie@epa.gov)

## Accessing SeqAPASS

For optimal SeqAPASS performance use **Chrome**

Access SeqAPASS using the following URL: <https://www.seqapass.epa.gov/seqapass/>

### Returning Users

Click “Login”

**New to SeqAPASS Version 7.0 (See [user guide](#) for more details)**

- Automatic filtering for Level 3 sequence alignment selections based on annotations
- Advanced users can request Level 4 access for protein structure generation and alignment to gather additional lines of evidence toward conservation
  - Conversion of aligned sequences from Level 1 to FASTA format for Level 4
  - Aided selection of protein sequences for submission to I-TASSER (<https://zhanggroup.org/I-TASSER/>) for creation of protein structural models
  - Ability to select protein structures from AlphaFold (<https://alphafold.ebi.ac.uk/>) and other sources using PDB formatted protein structures for structural alignment
  - Align structures to understand cross species protein conservation using TM-align (<https://zhanggroup.org/TM-align/>)
- Streamlined process for integration of Level 4 data for evaluating protein structures across species in the context of understanding conservation, extrapolating knowledge across species, and predicting chemical susceptibility

Log In to SeqAPASS Version 7.0

Welcome to SeqAPASS

[Login](#)

For optimal SeqAPASS performance use Chrome

Instructions to create a SeqAPASS account and login can be found [here](#).

[About SeqAPASS](#)

Select either “Login with EPA LAN User ID & Password” or “Login with Single Sign-On”.



Login with ...

User ID & Password

Single Sign-On

Login with User ID & Password

Login with Single Sign-On

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: [LaLone.Carlie@epa.gov](mailto:LaLone.Carlie@epa.gov)

### First time users

To request a username and password to access the SeqAPASS tool, select “[here](#)” below the login and follow the directions on the next page. The directions are different for the internal EPA user versus the external non-EPA user; however the user type does not limit access to the tool. Everyone that requests an account will be given one in a timely manner. Individual account allows users to store all previous SeqAPASS runs. Once the user has obtained their username, external users will select “Login with EPA LAN User ID and Password.”

#### EPA Users

1. Go to <https://waa.epa.gov> and login with your existing EPA LAN id and password.
2. Under the "Community Access" menu, select "Request Web Community Access"
3. Select the "SeqAPASS Users" community and click submit.
4. Return to the SeqAPASS login page to access SeqAPASS

#### External Users

1. Go to <https://waa.epa.gov> and click on the "Self Register" link.
2. Fill out the form using the following EPA Contact information:
  - o EPA Contact Name - *Carlie Lalone*
  - o EPA Contact's Email Address - *lalone.carlie@epa.gov*
  - o EPA Contact's Phone Number - *218-529-5038*
3. Select the "SeqAPASS Users" community from the dropdown menu at the bottom of the page.
4. Once you submit the form you will receive an email confirming your request and a follow-up email with your username once your account has been activated.

On the Log in screen the user will provide the necessary Login information:

EPA User: EPA LAN User ID & Password **or** Login with Windows Single Sign-On  
External User: Username and Password

Upon creating your password, login to SeqAPASS as described above for *Returning Users*. To change a password at any time, go to [waa.epa.gov](https://waa.epa.gov) and select “User Profile” to reset. The user will then use the new password to login.

### Messages from the SeqAPASS development team

Look for messages about planned version releases, data updates, and/or fixes to the SeqAPASS tool. These will occasionally be displayed below the SeqAPASS banner when the development team has information to share with SeqAPASS users.

**New to SeqAPASS Version 7.0 (See [user guide](#) for more details)**

- Automatic filtering for Level 3 sequence alignment selections based on annotations
- Advanced users can request Level 4 access for protein structure generation and alignment to gather additional lines of evidence toward conservation
  - Conversion of aligned sequences from Level 1 to FASTA format for Level 4
  - Aided selection of protein sequences for submission to iTASSER (<https://zhanggroup.org/iTASSER/>) for creation of protein structural models
  - Ability to select protein structures from AlphaFold (<https://alphafold.ebi.ac.uk/>) and other sources using PDB formatted protein structures for structural alignment
  - Align structures to understand cross species protein conservation using TM-align (<https://zhanggroup.org/TM-align/>)
  - Streamlined process for integration of Level 4 data for evaluating protein structures across species in the context of understanding conservation, extrapolating knowledge across species, and predicting chemical susceptibility

Log In to SeqAPASS

Version 7.0

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## SeqAPASS Home Tab

The “Home” tab indicates who is logged in to the tool (right-hand of the screen) and contains links to obtain information about the SeqAPASS tool (About SeqAPASS), including contact information for support and references to published articles describing the SeqAPASS tool and its applications. Other relevant references to databases and tools are also referenced. A link to the SeqAPASS User Guide can also be found on this page. To Submit a Comment/Question click on the “Submit Comment/Question” link to email the developer. “Log out” icon in upper right-hand corner of screen can be clicked at any time to log out. “Information” buttons are present throughout SeqAPASS to give the user additional information or instruction regarding features and functionality of the tool. “Exit” buttons are also present by each external (non-EPA) link that takes the user to a page *NOT* maintained by the EPA.

Welcome to SeqAPASS

Version 7.0

Logged in as: Maxwell Botz

SeqAPASS Home

[About SeqAPASS](#)  
[SeqAPASS User Guide](#)  
[Submit Comment/Question or Report a Problem](#)

**Welcome to the Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Tool**

The SeqAPASS tool is intended to be used to extrapolate chemical toxicity knowledge/data from one species to others. This extrapolation is termed cross species extrapolation and is based on the current understanding of a chemical-protein or protein-protein interaction in one species and the collection of lines of evidence indicating a similar interaction in another species is likely to occur. The results from SeqAPASS provide evidence of protein conservation to be used for predictions of chemical susceptibility or pathway conservation across the diversity of species. This is important as the majority of species will never be tested in the laboratory for toxicity through researchers and regulators are interested in understanding the potential for chemical impacts on all species that have potential for exposure in the environment.

Specifically, the tool takes advantage of existing knowledge of a chemical causing an effect through a particular protein in a particular species (usually a species known to be sensitive to a chemical) or a protein involved in an assay where a test species is represented. The protein from that species is queried in the SeqAPASS tool to generate lines of evidence for conservation in other species.

Note: Necessary input to SeqAPASS is the combination of a protein and an individual species

There are three levels to the SeqAPASS evaluation available to all users: Level 1 compares primary amino acid sequences, Level 2 compares functional domains, and Level 3 compares critical individual amino acids across species. Each level provides an additional line of evidence toward conservation and a prediction of susceptibility, which is equivalent to saying the protein is conserved in that species and the chemical is likely to interact similar to the query species. Users will choose to run Level 2 or Level 3 if there is enough knowledge available relative to functional domains and critical amino acids. There are information buttons integrated throughout the tool to guide users through this streamlined analysis as well as visualizations, summary reports, and strategic interoperability with other valuable resources for interpreting the results for cross species extrapolation.

**Demonstrated uses of SeqAPASS**

Academic, industry, and government researchers and decision-makers are using this tool to inform studies and decisions, including:

- Predicting chemical susceptibility across the diversity of species
- Defining how broadly data from toxicity studies may be extrapolated to other non-tested species (including threatened and endangered species)
- Defining the taxonomic domain of applicability for biological pathways, such as adverse outcome pathways
- Predicting bioaccumulation potential of chemicals known to bioaccumulate through interactions with proteins
- Generating research hypotheses for further laboratory study

**Advanced Users of SeqAPASS Level 4**

The Level 4 evaluation is intended for advanced users only and generates protein structural models to perform structural alignments for an additional line of evidence of conservation. Such results can also be exported for more advanced bioinformatics approaches like molecular docking, virtual screening, or molecular dynamic simulations.

**Want to learn more?**

Please feel free to reach out the SeqAPASS team with any questions or to request training [SeqAPASS.support@epa.gov](mailto:SeqAPASS.support@epa.gov)

**Presentations:**

Presenter: Dr. Carlie LaLone, US EPA  
Title: Bioinformatics for Cross Species Extrapolation  
Host: US EPA  
Link: <https://youtu.be/TaavV69vWg8>

Presenter: Dr. Carlie LaLone, US EPA  
Title: Bioinformatics is for Everyone: Applications to Challenges in Ecotoxicology  
Host: Eawag – aquatic research  
Link: [https://www.youtube.com/watch?v=STdEUV\\_d9M](https://www.youtube.com/watch?v=STdEUV_d9M)

## Request SeqAPASS Run Tab

Clicking the “Request SeqAPASS Run” tab opens a page to enter the query information necessary for a SeqAPASS run. Each section of the “Request SeqAPASS Run” will be described below:

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

Log out

Home | Request SeqAPASS Run | SeqAPASS Run Status | View SeqAPASS Reports | Settings

Request Level 1 SeqAPASS Run

Version 7.0

Logged in as: Maxwell Botz

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Identify a Protein Target

SeqAPASS is designed to predict cross species chemical susceptibility. Protein targets are often decided based on chemical, adverse outcome pathway (AOP), or high-throughput screening (HTS) assay target. Resources have been provided, as links, to aid the user in searching for appropriate protein targets and can be accessed by selecting the drop-downs found in the “Identify a Protein Target” box.

#### Identify a Protein Target

SeqAPASS is designed to predict cross species chemical susceptibility based on a protein molecular target. The following resources have been identified to guide the user to an appropriate protein target based on the chemical, adverse outcome pathway (AOP), or high-throughput screening (HTS) assay target of interest. Click the help buttons below for descriptions of how to find relevant protein target information from these resources.

All links will open in a new tab.

The following links exit the site [EXIT](#)

▼ **Pharmaceutical protein targets:**

- <https://www.drugbank.ca>
- <http://sitem.herts.ac.uk/aeru/vsdb/index.htm>
- [http://bidd.nus.edu.sg/group/cjtttd/TTD\\_HOME.asp](http://bidd.nus.edu.sg/group/cjtttd/TTD_HOME.asp)

▼ **Pesticides and other chemical protein targets:**

- <http://www.t3db.ca>

▼ **AOP chemical initiators:**

- <https://aopwiki.org>

▼ **ToxCast HTS results by chemical:**

- <https://comptox.epa.gov/dashboard>

### Select Search

There are two options for entering query information: “By Species” or “By Accession” (See radio buttons to the right of “Select Search”). Selecting “By Species” will allow the user to enter text and select from a dropdown list of species and then select a protein from any sequence available for that species in the NCBI protein database. Selecting “By Accession” allows the user to enter a NCBI protein accession.

#### Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) [Log out](#)

[Home](#) [Request SeqAPASS Run](#) [SeqAPASS Run Status](#) [View SeqAPASS Reports](#) [Settings](#)

Request Level 1 SeqAPASS Run Version 7.0 Logged in as: Maxwell Botz

Identify a Protein Target

Compare Primary Amino Acid Sequences

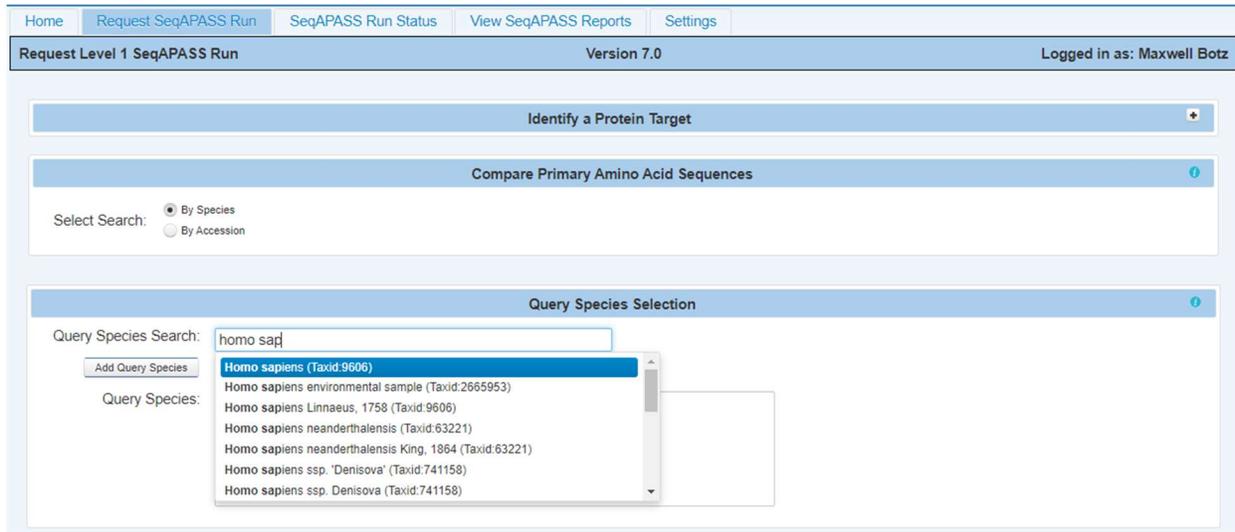
Select Search:  By Species  By Accession

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

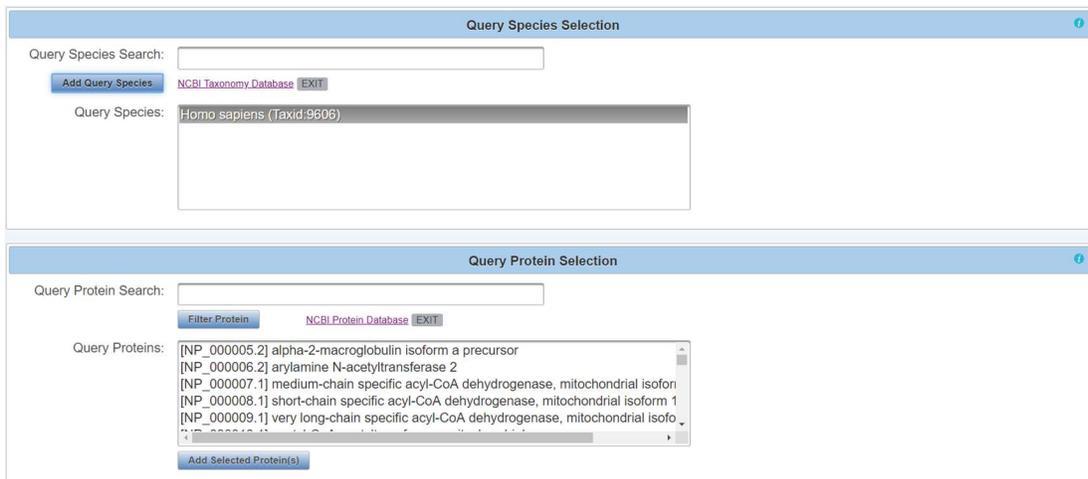
## Query “By Species”

Type the name of the query species of interest in the “Query Species Search” text box. The species common name, scientific name, or Taxid (ID number derived from the NCBI taxonomy database) may be typed into the search bar. This is the species you would like to compare all other species to. The search bar has an auto-complete function and will generate a list of species with corresponding Taxid. When text is typed into the search bar, the auto-complete function queries the database in the order of “starts with” then “contains.” If an integer is typed in the search bar the auto-complete function queries the database in the order of “Taxid”, “starts with”, then “contains.”



**Note:** The user can also use the NCBI taxonomy database to identify query species using the NCBI link on the right-hand side of the “Add Query Species” button.

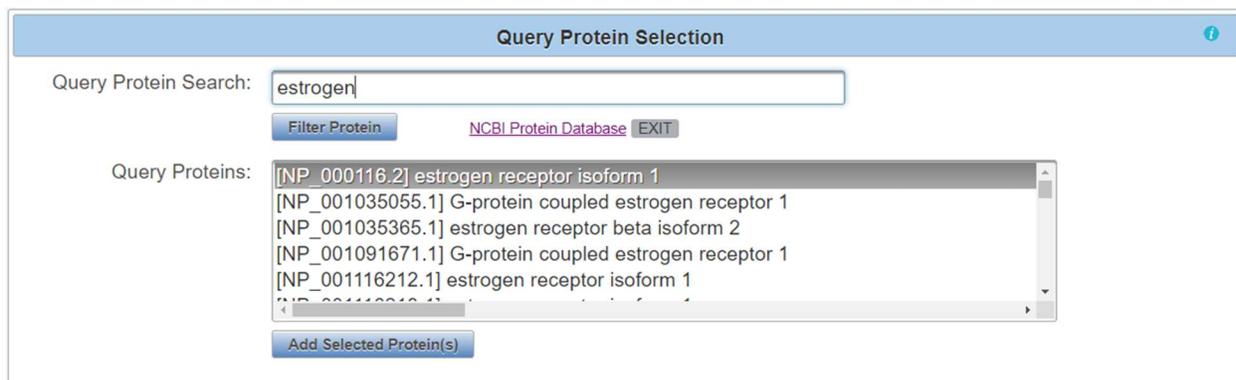
Select species of interest by clicking on the name in the drop-down box. Once species is selected, click “Add Query Species” button. This advances the species of interest to the “Query Species” box and fills the “Query Proteins” box with all available protein sequences for that species from the NCBI protein database (although the box only displays the initial 200 proteins/species based on lowest numerical accession number). The protein list includes the protein NCBI accession, protein name, and species scientific name.



## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

To filter the query protein list, type the query protein name or partial name in the “Query Protein Search” box and click the “Filter Protein” button. This action will filter the protein list in the “Query Proteins” box to only display proteins that contain the user defined text (this search query does not contain an autofill feature due to the filter feature). Proteins will be listed in alphabetical order based on NCBI accession. Example: typing “estrogen” retrieves all proteins that contain the word “estrogen” in the protein name (the user can scroll to identify proteins of interest).



Query Protein Selection

Query Protein Search:

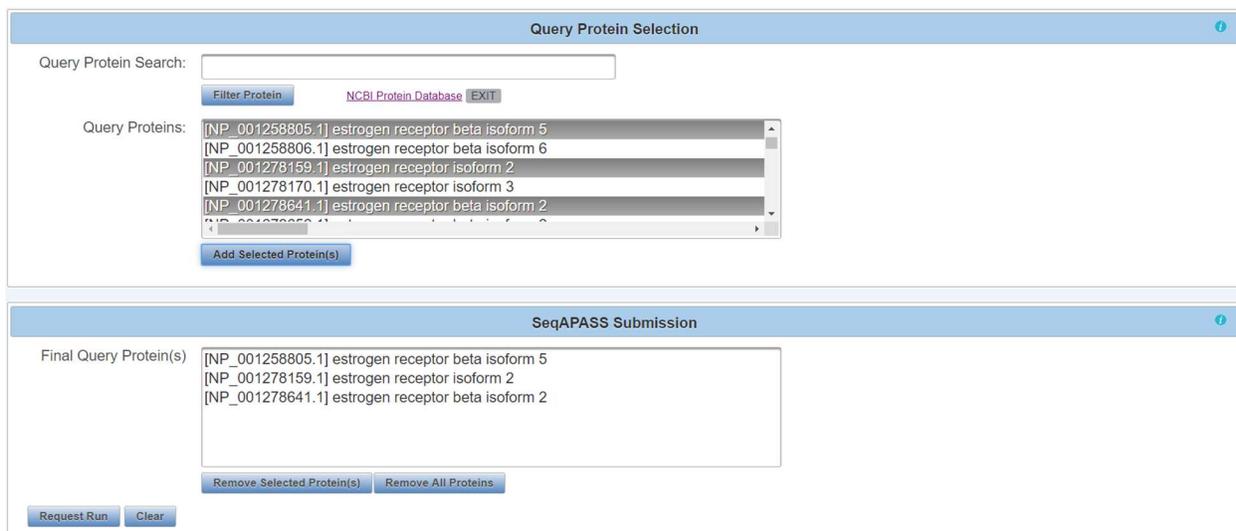
[NCBI Protein Database](#)

Query Proteins:

- [NP\_000116.2] estrogen receptor isoform 1
- [NP\_001035055.1] G-protein coupled estrogen receptor 1
- [NP\_001035365.1] estrogen receptor beta isoform 2
- [NP\_001091671.1] G-protein coupled estrogen receptor 1
- [NP\_001116212.1] estrogen receptor isoform 1

**Note:** To explore details associated with a protein of interest, click the “NCBI Protein Database” link to the right of the “Filter Protein” button to open NCBI proteins database (See [SeqAPASS Documentation](#) section of user guide for details about searching for query proteins using NCBI database).

Highlight the protein or proteins of interest (Ctrl left click to select multiple proteins) in the “Query Proteins” box and click “Add Selected Protein(s)” button. This moves the protein(s) of interest to the “Final Query Protein(s)” box. To remove proteins from the “Final Query Protein(s)” box highlight those to be removed and click the “Remove Selected Protein(s)” button. Select “Remove All Proteins” to discard all proteins from “Final Query Protein(s)” box. The clear button removes all information previously entered on the “Request SeqAPASS Run” page.



Query Protein Selection

Query Protein Search:

[NCBI Protein Database](#)

Query Proteins:

- [NP\_001258805.1] estrogen receptor beta isoform 5
- [NP\_001258806.1] estrogen receptor beta isoform 6
- [NP\_001278159.1] estrogen receptor isoform 2
- [NP\_001278170.1] estrogen receptor isoform 3
- [NP\_001278641.1] estrogen receptor beta isoform 2

SeqAPASS Submission

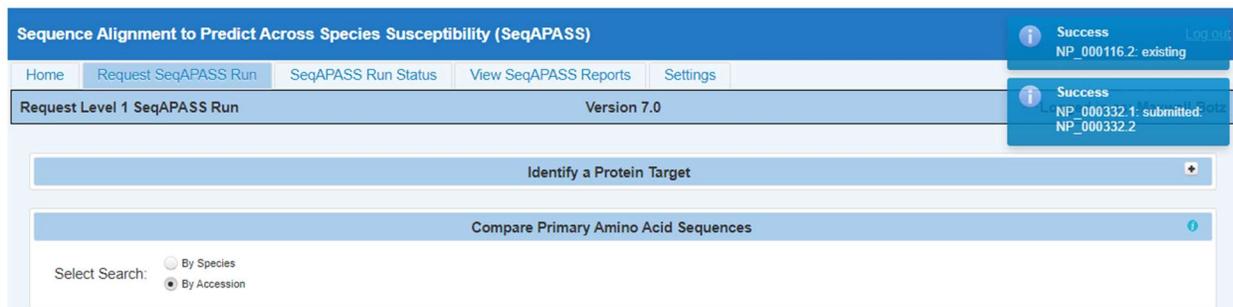
Final Query Protein(s)

- [NP\_001258805.1] estrogen receptor beta isoform 5
- [NP\_001278159.1] estrogen receptor isoform 2
- [NP\_001278641.1] estrogen receptor beta isoform 2

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

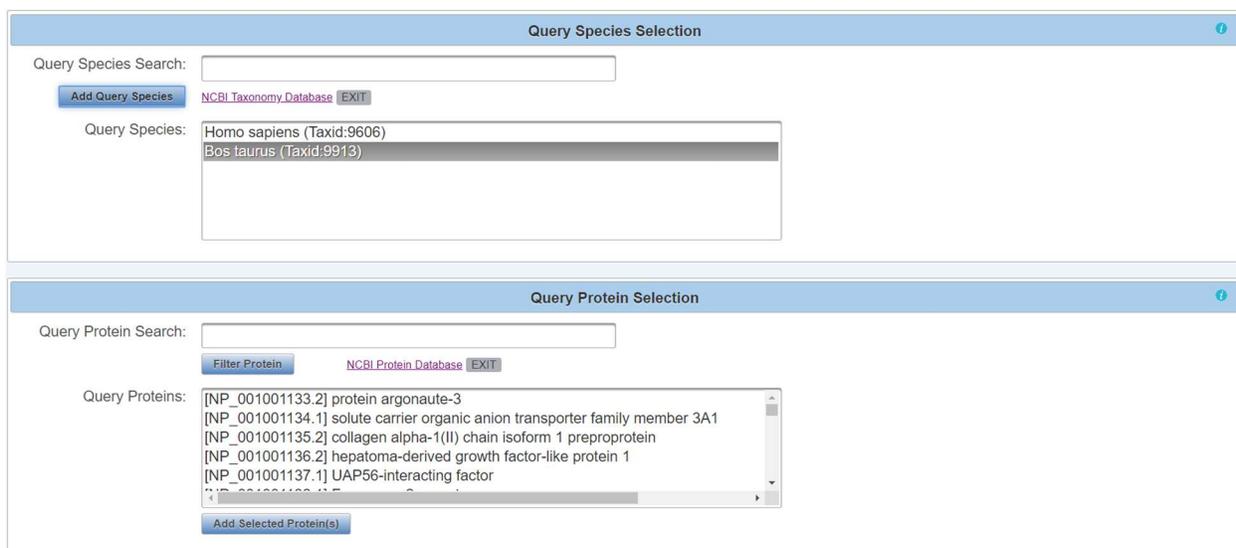
Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Once the user identifies the protein(s) to be queried, select “Request Run.” A message will briefly appear in upper right-hand corner of the screen for 10 seconds to alert the user of the request status.



Multiple proteins can be added to the final list for multiple SeqAPASS runs. If another query species is desired, return to “Query Species Search” to select the next species. Follow the process described above for selecting the proteins associated with this species. The proteins populated in the “Query Proteins” box will always be associated with the species highlighted in the “Query Species” box.

**Note:** In the current version of SeqAPASS, *PLEASE do not request more than 10 query proteins at a time* to avoid longer wait times for the completion of a run.



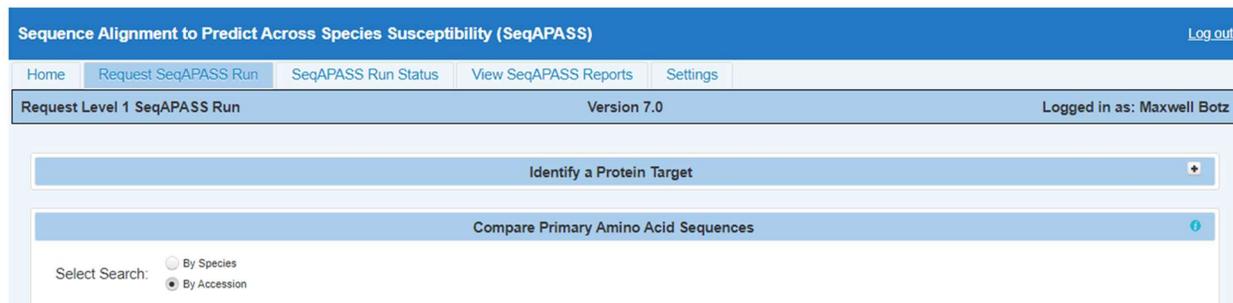
**Note:** A user may check the progress of the run by clicking on the “SeqAPASS Run Status” tab. (See **SeqAPASS Run Status** section of the user guide for more information)

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

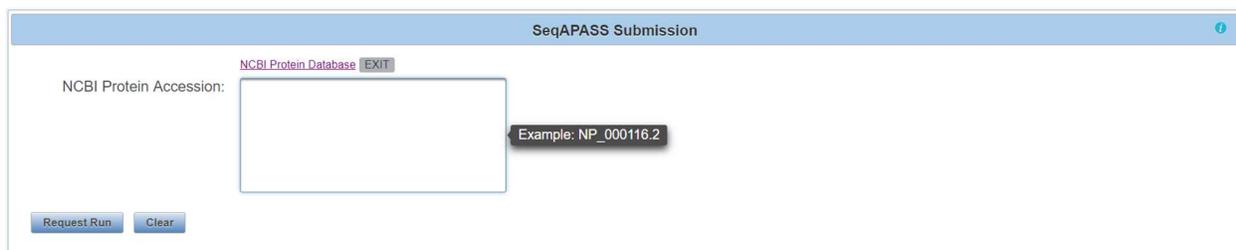
### Query “By Accession”

Users familiar with the NCBI database can utilize NCBI protein accessions (e.g., NP\_000116.2) to query the SeqAPASS tool. This is done by selecting the “By Accession” radio button to the right of the “Select Search” text on the “Request SeqAPASS Run” page.



Upon selecting the “By Accession” radio button, a new query page will be displayed. Type the NCBI protein accession (e.g., NP\_000116.2) for the protein of interest (this Accession comes from the NCBI protein database; See “**SeqAPASS Documentation**” for details) in the “NCBI Protein Accession” box. If desired, more than one NCBI Accession may be entered into the “NCBI Protein Accession” box by clicking the enter key after each additional NCBI Accession entry.

Upon clicking the “NCBI Protein Accession” text box, a pop-up message will appear in the middle of the text box, to provide an example for the proper format of Accessions to be entered.



**Note:** To avoid longer wait times for the completion of a run, in the current version of SeqAPASS, *please do not request more than 10 NCBI Accessions at a time.*

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Home Request SeqAPASS Run SeqAPASS Run Status View SeqAPASS Reports Settings

Request Level 1 SeqAPASS Run Version 7.0 Logged in as: Maxwell Botz

Identify a Protein Target

Compare Primary Amino Acid Sequences

Select Search:  By Species  By Accession

SeqAPASS Submission

NCBI Protein Accession:

Request Run Clear

After the NCBI accession(s) of interest have been typed in the “NCBI Protein Accession” box, click the “Request Run” button. To remove proteins from the “NCBI Protein Accession” box click the “Clear” button. A message will briefly appear in the upper right-hand corner of the screen to alert the user of their run request status.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

Home Request SeqAPASS Run SeqAPASS Run Status View SeqAPASS Reports Settings

Request Level 1 SeqAPASS Run Version 7.0

Identify a Protein Target

Compare Primary Amino Acid Sequences

Select Search:  By Species  By Accession

Success NP\_000116.2: existing

Success NP\_000332.1: submitted. NP\_000332.2

Note: All NCBI Accessions can include the version number (one digit after the decimal place, e.g., NP\_000116.2). Otherwise, if the version is not included, the most recent version of the accession will be queried automatically.

## SeqAPASS Run Status

Level 1 SeqAPASS (primary amino acid sequence comparisons) status is displayed as the default. The Accession in the column “Level 1 Query Accession” is that selected and queried by the user. For a query to finish it must display “complete” in the BLASTp column, 100% in the “Common Domains” column, and 100% in the “Ortholog Candidate” column. The “Common Domains” column displays the % completion for running Reverse Position Specific (RPS)-BLAST (Default E-value of  $\leq 0.01$ ) on the Accessions from the Level 1 Full Report. RPS-BLAST, and therefore “Common Domains” status, will take the longest to complete. The “Ortholog Candidate” column displays the % completion for running a reciprocal best hit BLAST evaluation for each hit sequence. The status for the “BLASTp” column is described as “started,” “analyzing,” or “complete.” If the user’s successfully submitted query has entered the run queue, the position of the submitted query in the queue will be indicated in the column (e.g., 2<sup>nd</sup> in queue). The “Common Domains” and “Ortholog Candidate” columns will also describe the position of the user’s submitted query in the run queue. Once the run has begun processing, the % completed for RPS-BLAST or reciprocal best hit BLAST, respectively, will be displayed. Please see example below:

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

SeqAPASS Run Status Version 7.0 Logged in as: Maxwell Botz

Level 1 Status  
 Level 2 Status  
 Level 3 Status  
 Level 4 Status

Refresh Data

SeqAPASS Level 1 Run Status

Search: Enter keyword

SeqAPASS Run Id	Data Version	User	Level 1 Query Accession	BLASTp	Common Domains	Ortholog Candidate	Start Date	Date Completed	SeqAPASS Run Duration
2419	7	Botz.Maxwell@epa.gov	NP_000332.2	started	1st in queue	1st in queue	2023 08 03 10:07:04	Not Finished	-
2419	7	Botz.Maxwell@epa.gov	NP_000116.2	complete	100%	100%	2023 08 03 10:07:04	2023 08 03 10:07:04	1 seconds
2418	7	lalone.carlie@epa.gov	NP_001296002.1	complete	100%	100%	2023 08 03 09:09:15	2023 08 03 09:09:15	1 seconds
2417	7	Jensen.Marissa@epa.gov	XP_039539628.1	complete	100%	100%	2023 08 02 15:56:55	2023 08 02 17:18:07	1 hour(s) 21 minute(s)
2416	7	lalone.carlie@epa.gov	P02766.1	complete	100%	100%	2023 08 02 14:45:42	2023 08 02 14:45:42	1 seconds
2415	7	lalone.carlie@epa.gov	P02766.1	complete	100%	100%	2023 08 02 14:44:48	2023 08 02 14:44:48	1 seconds
2414	7	Jensen.Marissa@epa.gov	NP_001296002.1	complete	100%	100%	2023 08 02 14:39:52	2023 08 02 14:39:52	1 seconds
2413	7	null	NP_001296002.1	complete	100%	100%	2023 08 02 14:24:00	2023 08 02 14:24:00	1 seconds
2412	7	Schumann.peter@epa.gov	NP_000116.2	complete	100%	100%	2023 08 02 14:20:01	2023 08 02 14:20:01	1 seconds
2411	7	null	NP_001296002.1	complete	100%	100%	2023 08 02 14:16:04	2023 08 02 14:16:04	1 seconds

(1 of 617) Download Table

The user can view the status of requested SeqAPASS runs. Each Run is assigned a unique “SeqAPASS Run Id.” A **Run** is considered a query that was requested either individually or as a batch in the “Request SeqAPASS Run” tab. The user can view run start and end dates/times, and the duration of the run. (See **Search, View, and Download Data Tables** section of user guide for more information). The “Data Version” column indicates which version of NCBI data is being used (See “About” page for details on Data Versions)

The user is also able to view the status of Level 2 (Functional domain(s)), Level 3 (individual amino acid residue alignments), and Level 4 (FASTA, I-TASSER, and TM-align results).

View Level 2 Status by selecting the radio button. Also, while viewing the page, the user can click the “Refresh Data” button to refresh the data. “Level 1 Query Accession” column displays the NCBI accession selected and queried by the user. Please see below:

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

SeqAPASS Run Status Version 7.0 Logged in as: Maxwell Botz

Level 1 Status  
 Level 2 Status  
 Level 3 Status  
 Level 4 Status

Refresh Data

SeqAPASS Level 2 Run Status

Search: Enter keyword

SeqAPASS Run Id	Data Version	User	Level 1 Query Accession	NCBI Accession	Domain Type	BLASTp	Start Date	Date Completed
5119	7	Jensen Marissa@epa.gov	NP_001296002.1	NP_001296002.1	lon_trans	complete	2023 08 02 14:48:01	2023 08 02 14:49:18
5118	7	lalone.carlie@epa.gov	P02766.1	P02766.1	TLP_Transthyretin	complete	2023 08 02 14:47:33	2023 08 02 14:55:31
5117	7	Boltz.Maxwell@epa.gov	NP_000917.3	NP_000917.3	NR_LBD	complete	2023 03 30 10:42:10	2023 03 30 10:44:15
5116	7	Boltz.Maxwell@epa.gov	NP_000116.2	NP_000116.2	NR_LBD	complete	2023 01 27 11:19:46	2023 01 27 11:20:19
5115	7	Boltz.Maxwell@epa.gov	NP_000129.3	NP_000129.3	TB	complete	2022 11 29 10:53:54	2022 11 29 10:53:58
5114	7	Boltz.Maxwell@epa.gov	NP_000130.1	NP_000130.1	CD20	complete	2022 11 29 10:49:16	2022 11 29 10:49:19
5113	7	Boltz.Maxwell@epa.gov	NP_000128.1	NP_000128.1	FAA_hydrolase	complete	2022 11 16 07:31:40	2022 11 16 07:32:24
5112	7	Boltz.Maxwell@epa.gov	AAA98797.1	AAA98797.1	Serum_albumin	complete	2022 11 03 12:00:13	2022 11 03 12:00:15
5111	7	Boltz.Maxwell@epa.gov	NP_000116.2	NP_000116.2	NR_DBD_like	complete	2022 11 01 10:58:33	2022 11 01 10:59:06
5110	7	Boltz.Maxwell@epa.gov	NP_000116.2	NP_000116.2	NR_DBD_PNR_like_2	complete	2022 10 31 14:26:56	2022 10 31 14:27:26

(1 of 413) Download Table:

View Level 3 Status by selecting the radio button. “Level 1 Query Accession” column displays the NCBI accession selected and queried by the user. The “Job Name” is the user defined name chosen to describe the Level 3 alignment. Also, while viewing the page, the user can click the “Refresh Data” button to refresh the data. Please see below:

SeqAPASS Run Status Version 7.0 Logged in as: Maxwell Botz

Level 1 Status  
 Level 2 Status  
 Level 3 Status  
 Level 4 Status

Refresh Data

SeqAPASS Level 3 Run Status

Search: Enter keyword

SeqAPASS Run Id	Data Version	User	Job Name	Level 1 Query Accession	Template Accession	COBALT	Start Date	Date Completed
1179	7	Buglewicz Dylan@epa.gov	Transthyretin test by Dyl	P02766.1	AAD14937.2	complete	2023 08 02 16:04:04	2023 08 02 16:05:45
1178	7	lalone.carlie@epa.gov	Test_1_CL_8_2_23	AAI32976.1	AAI32976.1	complete	2023 08 02 15:42:52	2023 08 02 15:44:49
1177	7	Jensen Marissa@epa.gov	test	NP_001434.1	NP_001434.1	complete	2023 08 02 15:42:47	2023 08 02 15:44:13
1176	7	lalone.carlie@epa.gov	Test_CL_8_2	AAI32976.1	AAI32976.1	complete	2023 08 02 15:30:26	2023 08 02 15:30:33
1175	7	Schumann peter@epa.gov	test_8_2	NP_000116.2	NP_000116.2	complete	2023 08 02 15:24:07	2023 08 02 15:25:31
1174	7	Jensen Marissa@epa.gov	8_2 daphnid and fish no	NP_001296002.1	NP_001296002.1	complete	2023 08 02 15:21:08	2023 08 02 15:21:18
1173	7	lalone.carlie@epa.gov	Test_1_6_20_23	NP_000116.2	NP_000116.2	complete	2023 06 20 15:31:16	2023 06 20 15:31:23
1172	7	lalone.carlie@epa.gov	Test_6_20	NP_000116.2	NP_000116.2	complete	2023 06 20 15:28:35	2023 06 20 15:29:29
1171	7	Boltz.Maxwell@epa.gov	Cypress Test	NP_000116.2	NP_000116.2	complete	2023 02 02 11:47:46	2023 02 02 11:47:58
1170	7	Boltz.Maxwell@epa.gov	Multiple Full Runs Test	NP_000116.2	NP_000116.2	complete	2022 11 30 08:16:35	2022 11 30 08:16:41

(1 of 115) Download Table:

View Level 4 Status by selecting the radio button. “Level 1 Query Accession” column displays the NCBI accession selected and queried by the user. The “Job Name” is the user defined name chosen to describe the Level 4 FASTA Prioritization. While viewing this page the user may click the “Refresh Data” button to refresh the data. This page will display the status of each component of the Level 4 jobs from FASTA, I-TASSER, and TM-align.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

The screenshot shows the SeqAPASS Level 4 Run Status and TM-align Status sections. The Level 4 Status is selected. The I-TASSER Status table lists various runs with columns for SeqAPASS Run ID, Data Version, User, Job Name, Level 1 Query Accession, Status, Date Completed, # Completed, FASTA, and I-TASSER details. The TM-align Status table lists runs with columns for TM-align Run ID, Level 4 Run ID(s), Data Version, User, Job Name(s), TM-align Query Accession, Query Accession Length, Status, Start Date, Date Completed, and Duration.

To return to previous tabs click on “Home,” “Request SeqAPASS Run,” or “SeqAPASS Run Status” tabs.

## View SeqAPASS Reports Tab

The “View SeqAPASS Reports” tab provides a table of completed SeqAPASS runs. From this page the user can choose to either “View Report” or “Save Report(s).”

The screenshot shows the SeqAPASS Reports page. It includes a navigation bar with tabs for Home, Request SeqAPASS Run, SeqAPASS Run Status, View SeqAPASS Reports, and Settings. The View SeqAPASS Reports tab is active. Below the navigation bar, there are buttons for Request Selected Report and Refresh Available Reports. A radio button for View Report is selected, and a Save Report(s) option is also visible.

The completed runs, by default, are listed in the order in which they were completed, with the most recent runs at the top. The table includes information for each run, such as SeqAPASS Run ID (unique for every run regardless of if it is the same protein/species combination ran twice), Data Version, Ortholog Count (number of orthologs detected from the aligned hit sequences in Level 1; see Detailed Documentation page 79), NCBI Accession, Query Protein Name, taxonomy information for the query species, and the date/time of run completion.

While viewing the page, the user can click the “Refresh Available Reports” button to refresh the table with additional completed runs. Partial protein sequences are highlighted in yellow as illustrated in the

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

example below. (See **Search, View, and Download Data Tables** section of user guide for more information).

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Log out

Home Request SeqAPASS Run SeqAPASS Run Status View SeqAPASS Reports Settings

SeqAPASS Reports Version 7.0 Logged in as: Maxwell Botz

Partial Protein Sequence

View Report  
 Save Report(s)

Available Reports

Search: Enter keyword

SeqAPASS Run Id	Data Version	Ortholog Count	Level 1 Query Accession	Query Protein Name	NCBI Taxonomy ID	Query Species Name	Query Common Name
<input checked="" type="radio"/> 2419	7	712	NP_000116.2	estrogen receptor isoform 1	9006	Homo sapiens	Human
<input type="radio"/> 2418	7	225	NP_001296002.1	ryanodine receptor	51655	Plutella xylostella	Diamondback moth
<input type="radio"/> 2417	7	942	XP_039539628.1	LOW QUALITY PROTEIN: ryanodine receptor 2	90988	Pimephales promelas	Fathead minnow
<input type="radio"/> 2416	7	12	P02786.1	RecName: Full=Transthyretin; AltName: Full=ATTR; AltName: F	9006	Homo sapiens	Human
<input type="radio"/> 2410	7	272	NP_001075232.1	multidrug resistance-associated protein 1	9796	Equus caballus	Horse
<input type="radio"/> 2409	7	126	NP_001434.1	fatty acid-binding protein, liver	9006	Homo sapiens	Human
<input type="radio"/> 2406	7	419	P04150.1	RecName: Full=Glucocorticoid receptor; Short=GR; AltName: F	9006	Homo sapiens	Human
<input type="radio"/> 2405	7	174	AAI32978.1	AR protein	9006	Homo sapiens	Human
<input type="radio"/> 2403	7	3	NP_000035.2	androgen receptor isoform 1	9006	Homo sapiens	Human
<input type="radio"/> 2401	7	137	NP_001353.4	thyroxine 5-deiodinase	9006	Homo sapiens	Human

(1 of 463) 1 2 3 4 5 6 7 8 9 10 10 Download Table:

## View Report

To select a completed run and view Level 1 data, select the corresponding radio button in the first column of the table and click “Request Selected Report.” This will open the Level 1 page to view the Level 1 data and to set up queries for Level 2, Level 3, and Level 4.

**Note:** The user **MUST** select a radio button **PRIOR** to clicking “Request Selected Report.” If the user fails to select a radio button and clicks “Request Selected Report” a Spinning Wheel will appear and disappear, and no completed run will be opened. Further, **there is no pop-up message** indicating that the user did not select a radio button.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Log out

Home Request SeqAPASS Run SeqAPASS Run Status View SeqAPASS Reports Settings

SeqAPASS Reports Version 7.0 Logged in as: Maxwell Botz

Partial Protein Sequence

View Report  
 Save Report(s)

Available Reports

Search: Enter keyword

SeqAPASS Run Id	Data Version	Ortholog Count	Level 1 Query Accession	Query Protein Name	NCBI Taxonomy ID	Query Species Name	Query Common Name
<input checked="" type="radio"/> 2419	7	712	NP_000116.2	estrogen receptor isoform 1	9006	Homo sapiens	Human
<input type="radio"/> 2418	7	225	NP_001296002.1	ryanodine receptor	51655	Plutella xylostella	Diamondback moth
<input type="radio"/> 2417	7	942	XP_039539628.1	LOW QUALITY PROTEIN: ryanodine receptor 2	90988	Pimephales promelas	Fathead minnow
<input type="radio"/> 2416	7	12	P02786.1	RecName: Full=Transthyretin; AltName: Full=ATTR; AltName: F	9006	Homo sapiens	Human
<input type="radio"/> 2410	7	272	NP_001075232.1	multidrug resistance-associated protein 1	9796	Equus caballus	Horse
<input type="radio"/> 2409	7	126	NP_001434.1	fatty acid-binding protein, liver	9006	Homo sapiens	Human
<input type="radio"/> 2406	7	419	P04150.1	RecName: Full=Glucocorticoid receptor; Short=GR; AltName: F	9006	Homo sapiens	Human
<input type="radio"/> 2405	7	174	AAI32978.1	AR protein	9006	Homo sapiens	Human
<input type="radio"/> 2403	7	3	NP_000035.2	androgen receptor isoform 1	9006	Homo sapiens	Human
<input type="radio"/> 2401	7	137	NP_001353.4	thyroxine 5-deiodinase	9006	Homo sapiens	Human

(1 of 463) 1 2 3 4 5 6 7 8 9 10 10 Download Table:

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Save Report(s)

To download completed Level 1, 2, and 3 data, select the “Save Report(s)” radio button. Upon doing so the user can select which accession(s) to download by clicking the checkbox in the first column of the table associated with desired accession and click “Save Selected Report(s).”

The screenshot shows the SeqAPASS Reports interface. At the top, there are navigation tabs: Home, Request SeqAPASS Run, SeqAPASS Run Status, View SeqAPASS Reports, and Settings. The current page is 'SeqAPASS Reports', version 7.0, and the user is logged in as Maxwell Botz. Below the navigation is a control panel with a radio button for 'Partial Protein Sequence' (checked), a 'Save Selected Report(s)' button, and a 'Refresh Available Reports' button. Below this is a table of 'Available Reports'. The table has columns: SeqAPASS Run Id, Data Version, Ortholog Count, Level 1 Query Accession, Query Protein Name, NCBI Taxonomy ID, Query Species Name, and Query Common Name. The table contains 10 rows of data, with the first row highlighted in blue. At the bottom of the table, there is a pagination control showing '(1 of 463)' and a dropdown menu for page size (1-10).

The user can also deselect data that is not wanted in the download by scrolling to the far right of the table and deselecting the checkboxes for the different levels of the SeqAPASS analysis. By default, all available data for the selected accession will be downloaded in a zip file.

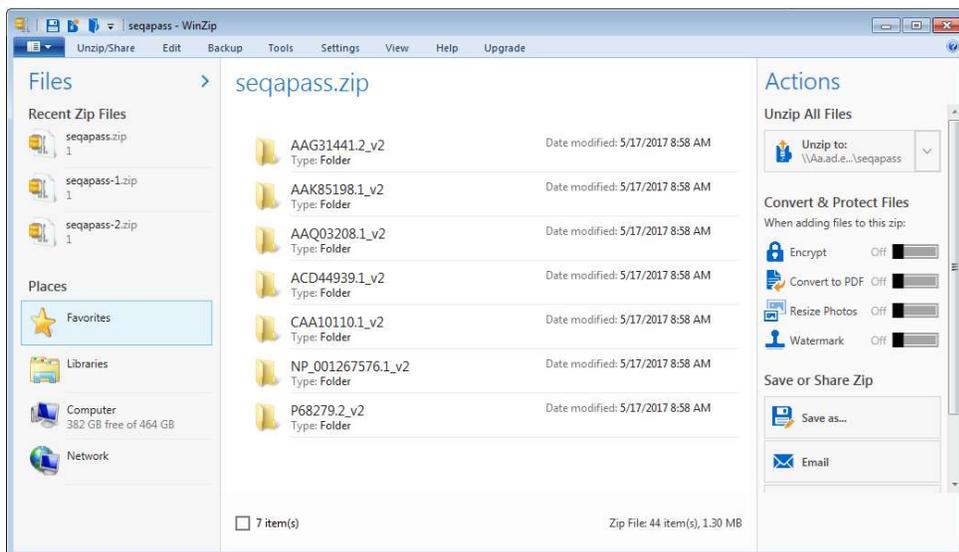
A WinZip file will be created for all the selected Reports.

The screenshot shows the SeqAPASS Reports interface with a 'Save As' dialog box open. The dialog box is titled 'Save As' and shows the file name 'seqapass.zip' and the save type 'WinZip File (\*.zip)'. The background shows the same table of available reports as the previous screenshot, with the 'Save Report(s)' radio button selected.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

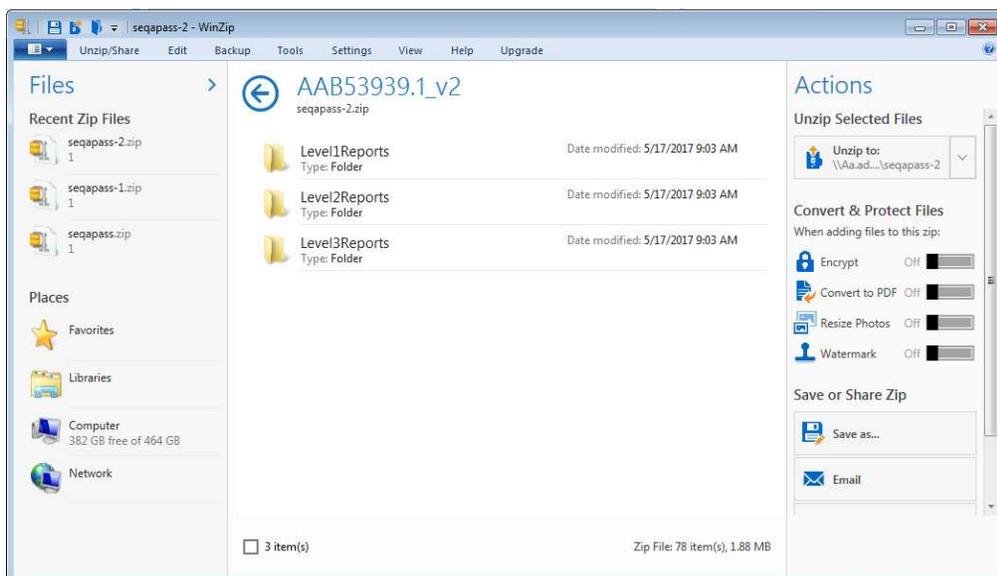
Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

A pop-up seqapass.zip file should appear with data files for each selected report. The naming convention is the NCBI Protein Accession and the Data Version (e.g., AAG31441.2\_v6).



By clicking on one of the Reports for a Protein Accession\_version, all available files for each Level of the SeqAPASS evaluation are available.

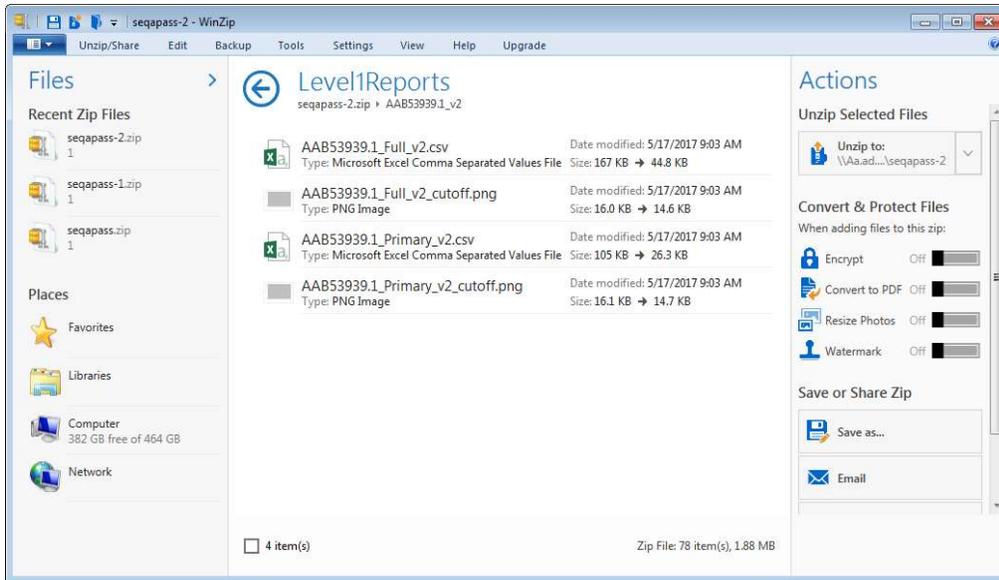
**Note:** This download includes default settings only. If susceptibility cut-off or any defaults were manipulated on Level 1 or 2 pages they will **NOT** be downloaded here and can **ONLY** be downloaded directly from the Level 1 or Level 2 page where the setting was manipulated by the user. Also, data visualizations can **ONLY** be downloaded from the Level 1 and 2 pages. They **DO NOT** populate in the zip file folders.



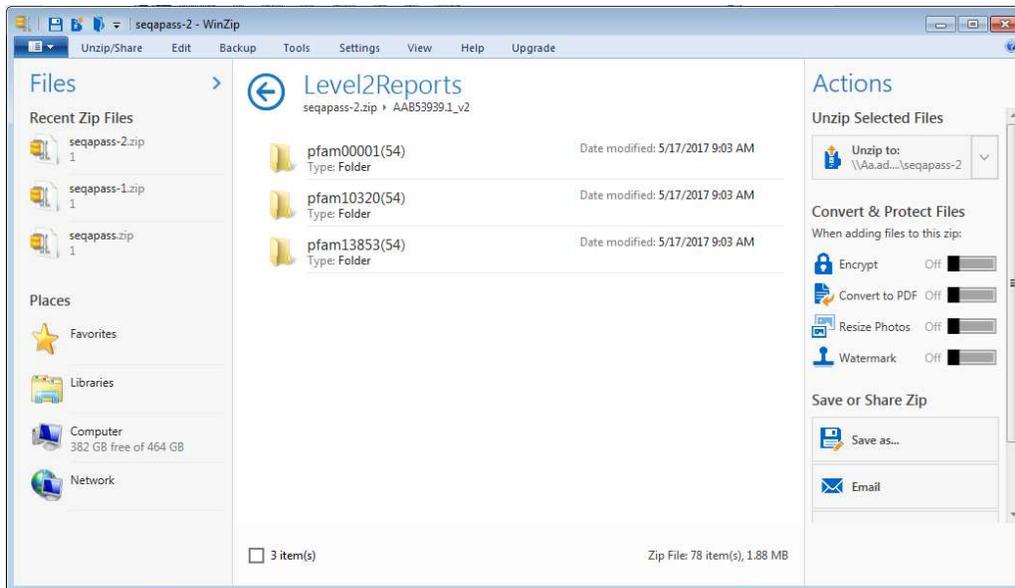
By selecting “Level1Reports”, both full and primary reports are available as csv files as well as a graphic of the density plot for determining the susceptibility cut-off.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov



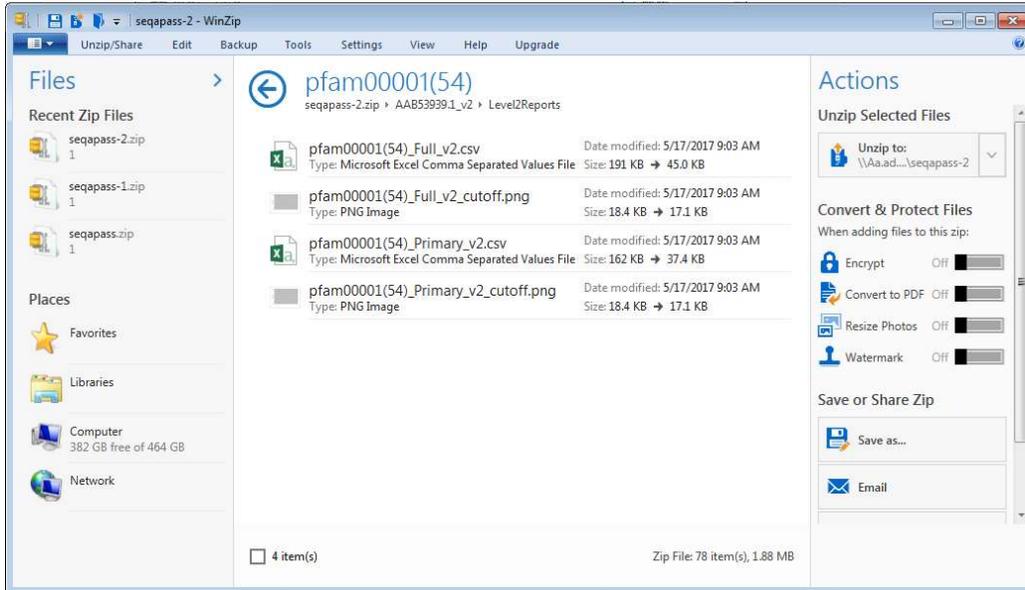
By selecting “Level2Reports”, all completed domain comparisons will be available and named by NCBI domain accession with the starting amino acid residue position for the domain (e.g., pfam00001(54)).



# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

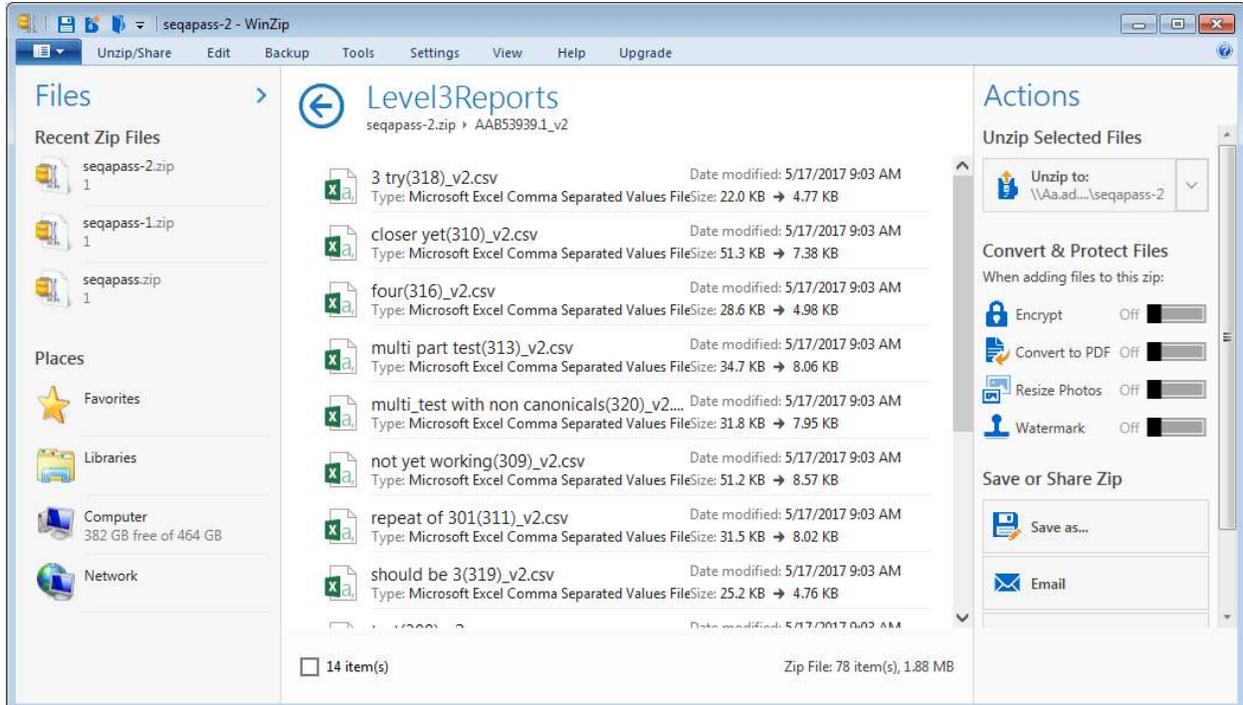
Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Upon selecting a domain file to view, both full and primary reports are available as csv files as well as a graphic of the density plot for determining the susceptibility cut-off.



By selecting “Level3Reports”, all user defined Level 3 alignments are available as csv.

Note: These csv files show the alignments across the entire sequence, not just those amino acid residues selected by the user.



# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Level 1: Primary Amino Acid Sequence Alignment

From the “View SeqAPASS Reports” tab, upon selecting a radio button and clicking “Request Selected Report” the Level 1 data will be displayed.

The “Level 1 Query Protein Information” box contains the SeqAPASS Run ID, Query Accession, Ortholog Count (# of hits identified as ortholog candidates to the query species protein sequence), NCBI Data updates (“Protein and Taxonomy Data:” displays the date that NCBI databases were downloaded and incorporated into the SeqAPASS database; BLAST Version: and Software Version: displays the version being used by the SeqAPASS tool for the selected data), Query Species, and Query Protein. Other information in this box will be described below.

The screenshot shows the SeqAPASS web interface. At the top, there is a blue header with the title "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)" and a "Log out" link. Below the header is a navigation bar with tabs: "Home", "Request SeqAPASS Run", "SeqAPASS Run Status", "View SeqAPASS Reports", and "Settings". The "View SeqAPASS Reports" tab is active. Below the navigation bar, there is a sub-header "SeqAPASS Reports" with "Version 7.0" and "Logged in as: Maxwell Botz". The main content area has tabs: "Main", "Level 1", and "DS Report". The "Level 1" tab is active. The "Level 1 Query Protein Information" section is displayed, showing the following details: "Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list." "SeqAPASS ID: 2419", "Query Accession: NP\_002118.2", "Ortholog Count: 712", "Protein and Taxonomy Data: 04/28/2021", "Query Species: Homo sapiens", "BLAST Version: 2.13.0", "Query Protein: estrogen receptor isoform 1", "Software Version: 6.1". Below this information are several interactive buttons: "Susceptibility Cut-off", "Primary Report Settings", "Visualization", "Level 2", "Level 3", "Level 4", "Refresh Level 4 Runs", and "Refresh Level 2 and 3 Runs".

The default table displayed at the bottom of the page is the “Primary Report”, which includes query protein information in the first row below the column titles, followed by hit proteins whose sequences aligned with the query protein. The hit proteins are ordered from the highest to lowest percent similarity (Maximum percent similarity = 100%). For each hit protein, Data version, NCBI Accession and species information is provided including the “Protein Count” which indicates the number of protein records per species in the NCBI protein database, taxonomic information (See **Primary Report Settings** section below in user guide for more detail on “Taxonomic Group” versus “Filtered Taxonomic Group” columns), and species names. Also included are the NCBI protein accession, protein name, BLASTp bitscore (describes overall quality of the alignment, See NCBI BLASTp tutorials), and percent similarity ( $[\text{hit bitscore}/\text{query bitscore}] * 100$ ). If the hit protein has been identified as an ortholog candidate (using reciprocal best hit blast method), it will be noted with a “Y” for yes or if not an ortholog candidate, a “N”, for no. If the hit protein is predicted to be susceptible according to the susceptibility cut-off criteria, that will also be noted with a “Y” for yes or alternatively an “N” for no. The date the analysis was completed is also identified. The data also includes a column describing the number of ortholog candidates identified using the reciprocal best hit BLAST method. The susceptibility cut-off is also listed in a column. The cut-off is determined through identifying local minimums in the density plot of the percent similarity values for the primary report data set and evaluation of ortholog candidates. Additionally, there is a column that identifies if the species is a Eukaryote noted with a “Y” for yes or alternatively an “N” for no. Links out to the NCBI Protein Database, NCBI Taxonomy Database, and ECOTOX Knowledgebase (specific to the data row) are embedded in the Level 1 data table for “NCBI Accession,” “Species Tax ID,” “Scientific Name,” “Protein Name”, and “ECOTOX” columns. (See **Search, View, and Download Data Tables** section of user guide for more information).

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Default highlights identify partial protein sequences, sequences with a bitscore higher than the query sequence and therefore percent similarity greater than 100% (commonly synthetic constructs), and when zero ortholog candidates are identified (in this case a user should consider a different query sequence or check the full report). Please see **Susceptibility Cutoff Box for Level 1** section of user guide for details when no orthologs are detected. Additionally, the default setting for the report shows only eukaryote data if a eukaryote is selected as the query protein, excluding prokaryote data from the table with the “Show Only Eukaryotes” checkbox checked. To view prokaryote data, deselect this checkbox. If a prokaryote is selected as the query protein, the default setting will include both eukaryote and prokaryote data and the “Show Only Eukaryotes” checkbox will not be selected. To limit the data to eukaryotes only, the user would check the “Show Only Eukaryotes” checkbox.

Columns in left side of table:

The screenshot shows the 'Level 1 Data - Primary' report interface. At the top, there are filter options: 'Primary Report' (selected), 'Full Report', 'Partial Hit Protein Sequence' (checked), 'Percent Similarity > 100%' (checked), 'Susceptible = Y, Ortholog Count = 1' (checked), and 'Show Only Eukaryotes' (checked). There are buttons for 'View Level 1 Summary Report' and 'Push Level 1 To DS Report'. Below the filters, there is a search bar and a table with the following columns: Data Version, NCBI Accession, Protein Count, Species Tax ID, Taxonomic Group, Filtered Taxonomic Group, Scientific Name, Common Name, and Protein Name. The table contains 10 rows of data for various species like Homo sapiens, Gorilla gorilla, Pan troglodytes, etc. At the bottom, there is a pagination bar showing '(1 of 137)' and a 'Download Table' button.

Columns in right side of table:

The screenshot shows the 'Level 1 Data - Primary' report interface, focusing on the right side of the table. The table has the following columns: Name, Protein Name, BLASTp Bitscore, Ortholog Candidate, Ortholog Count, Cut-off, Percent Similarity, Susceptibility Prediction, Analysis Completed, Eukaryote, and ECOTOX. The table contains 10 rows of data for various species like man, gorilla, panzee, wland gorilla, chimpanzee, orangutan, macaque, and mangabey. At the bottom, there is a pagination bar showing '(1 of 137)' and a 'Download Table' button.

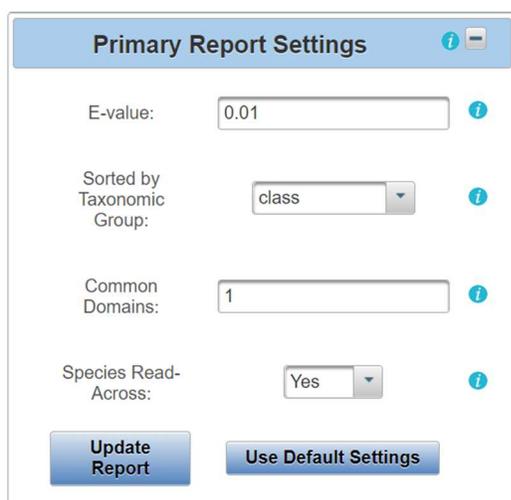
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Level 1: Primary Report Settings

#### *Default settings*

The “Primary Report Settings” drop down allows the user to view default settings on the table below and manipulate certain settings. “Primary Report Settings” are only available on the “Primary Report” display, not the “Full Report.” The default settings show data for hits whose E-value are  $\leq 0.01$  and have been identified to have  $\geq 1$  domain in common with the query sequence. The default setting for the “Sorted by Taxonomic Group” is “class,” therefore the “Filtered Taxonomic Group” column in the table is set to identify and report the taxonomic lineage of “class” from the NCBI Taxonomy Database. However, if class is not identified in the NCBI Taxonomic Hierarchy associated with the hit accession, then the algorithm will report the next available Taxonomic Group moving from class to subclass, to superorder, to order, to suborder, to superfamily, to family, to subfamily, to genus. Finally, the susceptibility predictions are set by using species read-across. (Please view **Documentation** Section of the User Guide for details on Read-Across settings). Briefly, Species Read-Across is used to set the susceptibility prediction, where all ortholog candidates are Susceptible = Y; all species listed above the susceptibility cut-off are Susceptible = Y; all species below the cut-off from the same taxonomic group of one or more species above the cut-off are Susceptible = Y; and those below the cut-off that are not ortholog candidates and do not belong to a taxonomic group above the cut-off are Susceptible = N.



The screenshot shows a window titled "Primary Report Settings" with a blue header. Inside, there are four settings, each with a text input or dropdown and an information icon (i):

- E-value: 0.01
- Sorted by Taxonomic Group: class
- Common Domains: 1
- Species Read-Across: Yes

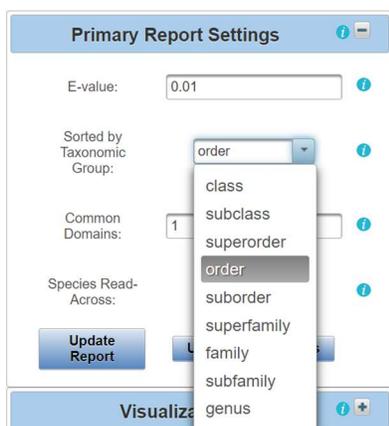
At the bottom, there are two buttons: "Update Report" and "Use Default Settings".

#### *Changing Default Settings*

The “E-value” and “Common Domains” settings can be manipulated by the user by entering the desired E-value or number of Common Domains in the respective text boxes and clicking “Update Report.” The table and data visualization will automatically be updated after a few seconds. The user may choose to change the level of the taxonomic hierarchy that is used for the susceptibility prediction. From the “Sorted by Taxonomic Group” dropdown the user may choose to display a different taxonomic group in the “Filtered Taxonomic Group” column of the data table.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov



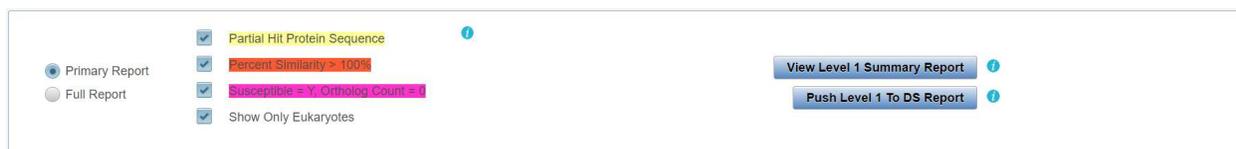
If the user chooses “order” for example, the “Filtered Taxonomic Group” column in the data table will report the taxonomic lineage of “order” from the NCBI Taxonomy Database and all species read-across for the susceptibility prediction will be based on order instead of class. The data visualization will also update. As described previously, if order is not identified in the NCBI Taxonomic Hierarchy associated with the hit accession, then the algorithm will report the next available taxonomic group moving from suborder, to superfamily, to family, to subfamily, to genus. Upon selecting the taxonomic group from the dropdown and clicking “Update Report,” the Level 1 Data for the Primary Report will update to the selected taxonomic level.

Search: <input type="text" value="Enter keyword"/>									
<input type="checkbox"/>	Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	
<input type="checkbox"/>	6	<a href="#">NP_000116.2</a>	2603582	<a href="#">9606</a>	Mammalia	Primates	<a href="#">Homo sapiens</a>	Human	
<input type="checkbox"/>	6	<a href="#">ABY64717.1</a>	1708	<a href="#">9593</a>	Mammalia	Primates	<a href="#">Gorilla gorilla</a>	Western gorilla	
<input type="checkbox"/>	6	<a href="#">XP_003311596.1</a>	171683	<a href="#">9598</a>	Mammalia	Primates	<a href="#">Pan troglodytes</a>	Chimpanzee	
<input type="checkbox"/>	6	<a href="#">XP_030868114.1</a>	52137	<a href="#">9595</a>	Mammalia	Primates	<a href="#">Gorilla gorilla gorilla</a>	Western lowland gorilla	
<input type="checkbox"/>	6	<a href="#">XP_003811544.1</a>	71982	<a href="#">9597</a>	Mammalia	Primates	<a href="#">Pan paniscus</a>	Pygmy chimpanzee	
<input type="checkbox"/>	6	<a href="#">ABY64718.1</a>	1609	<a href="#">9600</a>	Mammalia	Primates	<a href="#">Pongo pygmaeus</a>	Bornean orangutan	
<input type="checkbox"/>	6	<a href="#">XP_002817538.1</a>	141069	<a href="#">9601</a>	Mammalia	Primates	<a href="#">Pongo abelli</a>	Sumatran orangutan	
<input type="checkbox"/>	6	<a href="#">XP_011751932.1</a>	68712	<a href="#">9545</a>	Mammalia	Primates	<a href="#">Macaca nemestrina</a>	Pig-tailed macaque	
<input type="checkbox"/>	6	<a href="#">XP_011922091.1</a>	66421	<a href="#">9531</a>	Mammalia	Primates	<a href="#">Cercocebus atys</a>	Sooty mangabey	
<input type="checkbox"/>	6	<a href="#">XP_005552209.1</a>	98680	<a href="#">9541</a>	Mammalia	Primates	<a href="#">Macaca fascicularis</a>	Crab-eating macaque	

(1 of 137) [Navigation buttons] 10 Download Table:

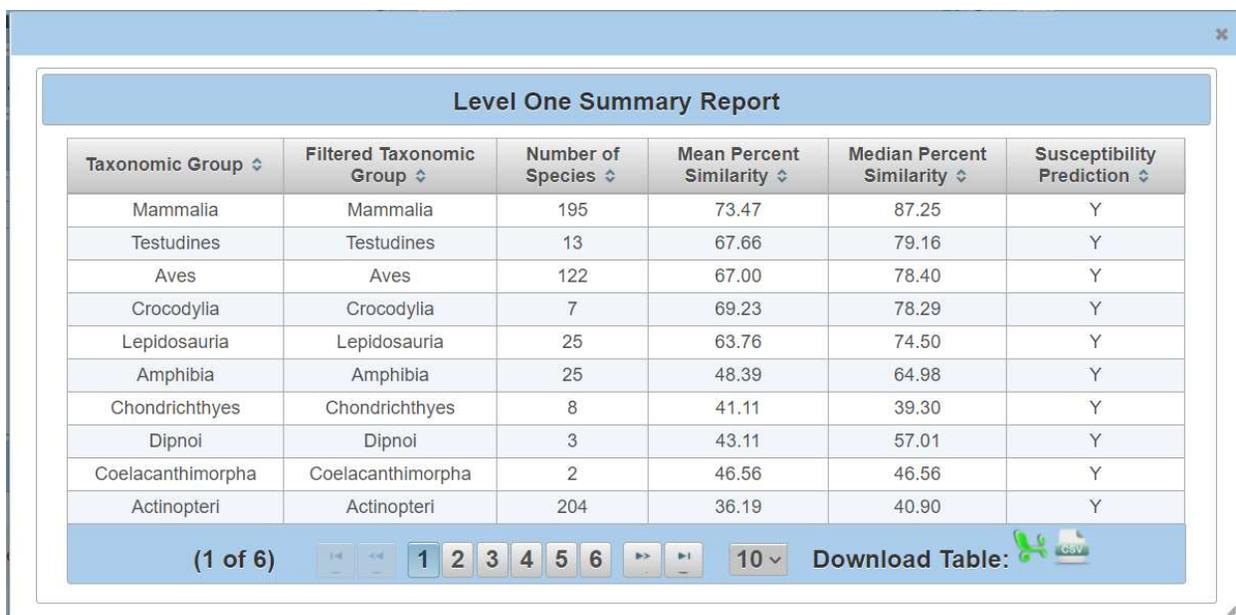
## Level One Summary Report

The user can view a summary of the data for each taxonomic group by clicking on the “View Level 1 Summary Report” button. The data includes, number of species, mean percent similarity, median percent similarity and susceptibility prediction. This data can also be downloaded.



## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

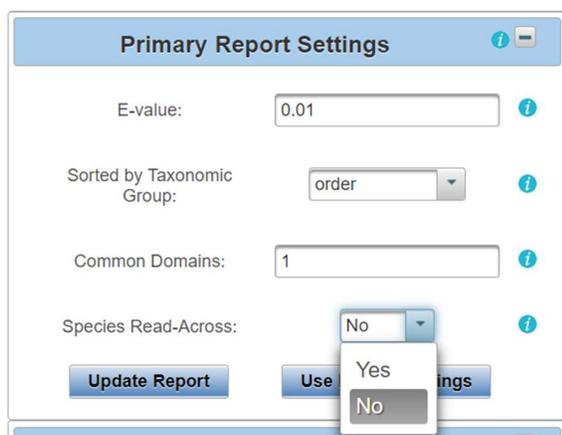
Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov



Taxonomic Group	Filtered Taxonomic Group	Number of Species	Mean Percent Similarity	Median Percent Similarity	Susceptibility Prediction
Mammalia	Mammalia	195	73.47	87.25	Y
Testudines	Testudines	13	67.66	79.16	Y
Aves	Aves	122	67.00	78.40	Y
Crocodylia	Crocodylia	7	69.23	78.29	Y
Lepidosauria	Lepidosauria	25	63.76	74.50	Y
Amphibia	Amphibia	25	48.39	64.98	Y
Chondrichthyes	Chondrichthyes	8	41.11	39.30	Y
Dipnoi	Dipnoi	3	43.11	57.01	Y
Coelacanthimorpha	Coelacanthimorpha	2	46.56	46.56	Y
Actinopteri	Actinopteri	204	36.19	40.90	Y

(1 of 6) [Navigation buttons] 10 [Download Table: CSV]

The user may also choose to turn species read-across off, by using the “Species Read-Across” drop-down and selecting “No” and clicking “Update Report.” When “No” is selected, the susceptibility predictions will only be “Y” in the table below if Percent Similarity is above the Cut-off or if the hit is identified as an Ortholog Candidate, yes or “Y.” Any hit below the cut-off will yield a susceptibility prediction of no or “N.”



**Primary Report Settings**

E-value: 0.01

Sorted by Taxonomic Group: order

Common Domains: 1

Species Read-Across: No (dropdown menu open with Yes and No options)

[Update Report] [Use] [Settings]

The user can select the “Full Report” on the “Level 1” page, which includes the same information as the “Primary Report” and additional information pertaining to the alignment of the protein sequence using BLASTp. Additional information includes the number of amino acid residues in the sequence (Hit Length), the number of exact matching amino acids between the hit and query sequence (Identity), the number of exact and similar matches in amino acids between the hit and the query sequence (Positives), the expect value (E-value) describing the number of different alignments expected to occur in the database search by chance, and the conserved domain count. The conserved domain count identifies all domains associated with the query protein in the NCBI conserved domains database (Specific hits, Non-specific hits, Superfamilies, and Multi-domains; See NCBI conserved domains database for details). SeqAPASS algorithms record the query sequence coverage of each curated domain and compares that coverage to that of the hit sequence. If the hit sequence covers the curated domain greater than or equal to

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

the query sequence, then the domain is considered a common domain between the hit and query. The number of common domains comparing each hit sequence to the query sequence are summed and reported. This column displays “0” when the hit protein and query protein do not have any common domains. (See **Search, View, and Download Data Tables** section of user guide for more information). The user can also download the currently applied report settings by selecting the “Download Current Level 1 Report Settings.” This csv allows the user to track which settings were used or changed by the user when downloading a data table.

The screenshot displays the 'Level 1 Data - Full' report interface. At the top, there are filter options: 'Primary Report' (selected), 'Full Report', 'Partial Hit Protein Sequence', 'Percent Similarity > 100%', 'Susceptible = Y, Ortholog Count > 1', and 'Show Only Eukaryotes'. Action buttons include 'View Level 1 Summary Report' and 'Push Level 1 To DS Report'. Below the filters is a search bar with the text 'Search: Enter keyword'. The main table has 15 columns: Hit Length, Identity, Positives, Evaluate, BLASTp Bitscore, Ortholog Candidate, Ortholog Count, Cut-off, Common Domain Count, Percent Similarity, Susceptibility Prediction, Analysis Completed, Eukaryote, and ECOTOX. The table contains 13 rows of data. At the bottom, there is a pagination bar showing '(1 of 139)' and a 'Download Table' button.

**Note:** SeqAPASS v2.0 and newer parse the BLASTp query and hit accessions to identify all the species/accessions from identical proteins. Therefore, if a hit sequence represents multiple species, all species with the identical sequence will be found in the data tables for Level 1. To determine which sequence/species was identified from BLASTp as a hit and which sequence/species was parsed from the identical sequence, view the “Full Report” for Level, column “Identical Protein,” where “N” is indicative of the original hit sequence and “Y” is the parsed sequence.

	A	B
1	Level 1 Report Settings	
2		
3		
4	Analysis TimeStamp	2019 05 16 11:04:08
5	SeqAPASS version	3.2
6	Query Species	Homo sapiens
7	Query Protein	estrogen receptor isoform 1
8	Query Accession	NP_000116.2
9	Ortholog Count	348
10	L1 Cutoff	Default
11	L1 Cutoff Value	33.93221513
12	E-value	0.01
13	Sorted by Taxonomic Group	CLASS
14	Common Domains	1
15	Species Read Across	Y
16	Show Only Eukaryotes	Checked
17	Report	Primary

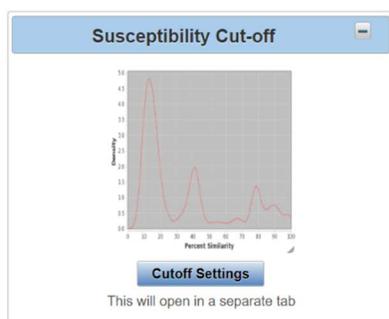
When downloading the current Level 1 report settings, the following information will be present in the csv file. If the user decides to change the default settings, the csv file can be utilized for quick information if the SeqAPASS page is no longer open.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

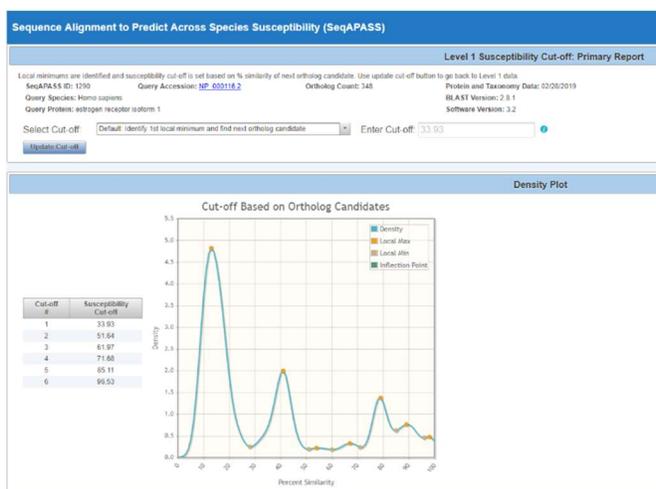
Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Susceptibility Cutoff Box for Level 1

The susceptibility prediction is determined by identifying ortholog candidates, sequences above a defined susceptibility cutoff, or by identifying those species below the susceptibility cut-off from an organism class above the susceptibility cutoff. The default susceptibility cut-off is set by plotting the distribution of percent similarities calculated for each hit protein. From this plot, the critical points are identified, and the local minimums and maximums reported. Using the ortholog candidate data, a susceptibility cut-off is automatically determined by identifying the first ortholog candidate at an equal or higher percent similarity than the first local minimum. The user can view this graph by clicking the “Cutoff Settings” button in the “Susceptibility Cut-off” box, which will open a new tab in the web browser. The “Select Cut-Off” drop-down can allow the user to select between the default cut-off, the 2<sup>nd</sup> local minimum or a user defined cut-off. The 2<sup>nd</sup> susceptibility cut-off is identified in the density plot by finding the 1<sup>st</sup> ortholog candidate at an equal or higher percent similarity to that of the 2<sup>nd</sup> local minimum. Upon selecting the User defined cut-off from the dropdown, the user can view and closely examine the density plot and manipulate the cut-off. The “Enter Cut-off” text box becomes active and the user can enter a number 1-100. To update the cut-off in the Level 1 data report and/or close the cutoff tab and return to the Level 1 page, click “Update Cut-off” button.



**Note:** The user should have a justification for changing the susceptibility cut-off, either based on evaluation of Ortholog cutoffs in the data visualization or from empirical evidence.



All potential susceptibility cut-offs generated by the data distribution and ortholog candidate identification are reported in the table with columns “Cut-off #” and “Susceptibility Cut-off”. The user can use these numbers to define a cut-off if empirical evidence suggests that the “Default” or “2<sup>nd</sup> minimum” are not supported.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

*No Orthologs Detected*

Level 1 Query Protein Information

Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 2354      Query Accession: [JLV\\_A](#) EXIT      Ortholog Count: 0      Protein and Taxonomy Data: 04/28/2021

Query Species: *Amaranthus caudatus*      BLAST Version: 2.13.0

Query Protein: Crystal Structure Of Amaranthus Caudatus Agglutinin      Software Version: 8.1

Susceptibility Cut-off +

Primary Report Settings +

Visualization +

Level 2 +

Level 3 +

Refresh Level 2 and 3 Runs

Level 4 +

Refresh Level 4 Runs

Partial Hit Protein Sequence

Percent Similarity > 100%

Susceptible = 1; Ortholog Count = 0

Show Only Eukaryotes

View Level 1 Summary Report

Push Level 1 To DS Report

Level 1 Data - Primary

The following links exit the site: [EXIT](#)

Download Current Level 1 Report Settings

ECOTOX Widget

Search: Enter keyword

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name
7	<a href="#">JLV_A</a>	152	<a href="#">2587</a>	Magnoliopsida	Magnoliopsida	<i>Amaranthus caudatus</i>	Amaranth	<i>Crystal Structure Of Amaranthus Caudatus Agglutinin</i>
7	<a href="#">AAD33922.1</a>	234	<a href="#">28502</a>	Magnoliopsida	Magnoliopsida	<i>Amaranthus hypochondriacus</i>	Grain amaranth	<i>agglutinin</i>
7	<a href="#">PSS21891.1</a>	32103	<a href="#">1592841</a>	Magnoliopsida	Magnoliopsida	<i>Azadirachta indica var. indica</i>	Golden kiewfruit	<i>Natterin-3 like</i>
7	<a href="#">XP_021894283.1</a>	27185	<a href="#">3549</a>	Magnoliopsida	Magnoliopsida	<i>Carica papaya</i>	Papaya	<i>uncharacterized protein LOC110311942</i>
7	<a href="#">GFV94439.1</a>	52432	<a href="#">168718</a>	Magnoliopsida	Magnoliopsida	<i>Azadirachta indica</i>	Nashi-kazura	<i>hypothetical protein Acc_D9C0026350</i>
7	<a href="#">XP_022953531.1</a>	43968	<a href="#">3052</a>	Magnoliopsida	Magnoliopsida	<i>Cucurbita moschata</i>	Crookneck pumpkin	<i>uncharacterized protein LOC111488051</i>
7	<a href="#">XP_022952244.1</a>	49142	<a href="#">3051</a>	Magnoliopsida	Magnoliopsida	<i>Cucurbita maxima</i>	Winter squash	<i>uncharacterized protein LOC111488923</i>
7	<a href="#">XP_008490021.1</a>	30498	<a href="#">3059</a>	Magnoliopsida	Magnoliopsida	<i>Cucumis melo</i>	Muskmelon	<i>PREDICTED: uncharacterized protein LOC103498998</i>
7	<a href="#">XP_023549484.1</a>	43520	<a href="#">3684</a>	Magnoliopsida	Magnoliopsida	<i>Cucurbita pepo subsp. pepo</i>	Vegetable marrow	<i>uncharacterized protein LOC111807833</i>
7	<a href="#">ZAA0038925.1</a>	73944	<a href="#">119495</a>	Magnoliopsida	Magnoliopsida	<i>Cucumis melo var. melonella</i>	Muskmelon	<i>uncharacterized protein EBC27_scaf001223002095</i>

(1 of 3)      Download Table:

If no orthologs are detected from reciprocal best hit blast analysis, the “Ortholog Count” will be “0” at the top of the “Level 1 Query Protein Information” page. The cutoff will be set by the local minimums only, therefore the susceptibility prediction will NOT take into account ortholog candidates. ***It is recommended that the user checks the full report for ortholog candidates or identifies a different query sequence for the susceptibility predictions.*** Here, the susceptibility predictions will be highlighted in dark pink in the Level 1 data table to indicate that 0 orthologs were detected and the susceptibility cutoff was determined from plotting the distribution of percent similarities and identifying the local minimums.

Level 1 Query Protein Information

Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 1299      Query Accession: [APO40848.1](#) EXIT      Ortholog Count: 0      Protein and Taxonomy Data: 02/28/2019

Query Species: *Poa annua*      BLAST Version: 2.8.1

Query Protein: PsbA, partial (plastid)      Software Version: 3.2

Note: De-select the “Show Only Eukaryotes” checkbox to see if prokaryotes were identified as orthologs.

By clicking on the “Cutoff Settings” button when no orthologs are detected, the “Cut-off #” and “Susceptibility Cut-off” columns will report only the local minimum values.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

### Level 1 Susceptibility Cut-off: Primary Report

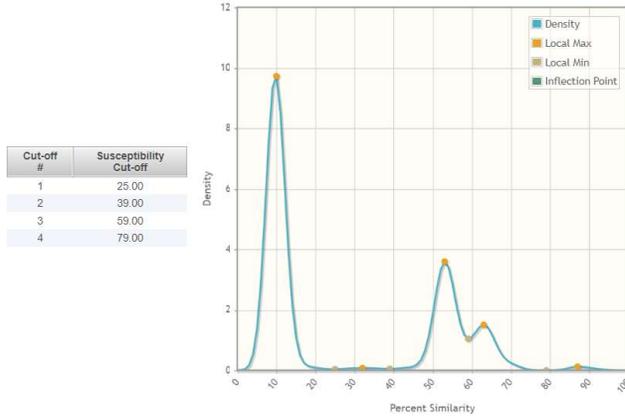
Local minima are identified and susceptibility cut-off is set based on % similarity of next ortholog candidate. Use update cut-off button to go back to Level 1 data.

SeqAPASS ID: 2297    Query Accession: [CAC38767.1](#)    Ortholog Count: 0    Protein and Taxonomy Data: 04/28/2021  
 Query Species: Pimephales promelas    BLAST Version: 2.11.0  
 Query Protein: cytochrome P450 aromatase    Software Version: 5.1

Select Cut-off:  Enter Cut-off:

### Density Plot

Cut-off Based on Ortholog Candidates



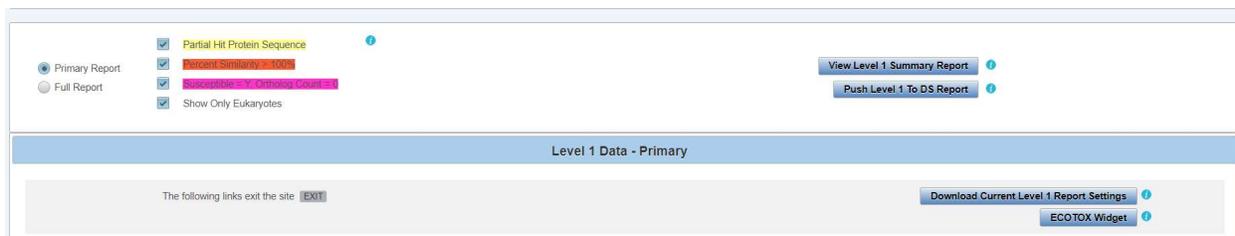
From the “Level 1” page the user can return to the list of completed SeqAPASS runs by clicking the “Main” button on the upper left-hand side of the “Level 1 Query Protein Information” page.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

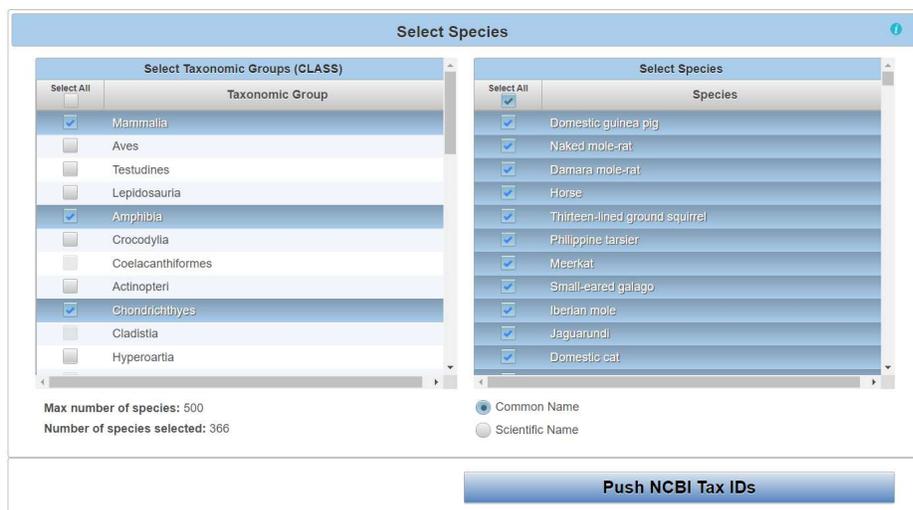
### ECOTOX Widget

The ECOTOX widget gives the user the option to create a species and chemical filter that will link out to ECOTOX. The widget allows for rapid access of curated empirical toxicity data from the ECOTOXicology (ECOTOX) Knowledgebase (<https://cfpub.epa.gov/ecotox/>) that can be compared to sequence-based predictions of chemical susceptibility from SeqAPASS results. In the “Level 1 Data” table header, the “ECOTOX Widget” button can be clicked and will open a widget that is populated with all the taxonomic groups and species from the Level 1 Data table.



### Select Species

Taxonomic groups that are present within the “Select Species” section of the ECOTOX widget are those found in the Level 1 Data table and Boxplot. Default settings auto select those taxonomic groups and species in common with ECOTOX. The user can select/deselect taxonomic groups of interest in the “Select Taxonomic Groups (CLASS)” box. Additionally, species can be selected/deselected in the “Select Species” box. Taxonomic groups and Species whose selection box is displayed greyed out, are not found in ECOTOX. The maximum number of species that can be pushed to the ECOTOX filter is 500. (Note: common species include those that are in the ECOTOX database, which does NOT mean they have toxicity records associated with them in ECOTOX.). Upon selecting species for comparison in ECOTOX, the user clicks on the “Push NCBI Tax IDs” button to advance to the “Select Chemicals” feature of the widget.



## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Select Chemicals (Optional)

The “Select Chemicals” feature of the ECOTOX widget is optional and can be skipped by selecting the “Open in ECOTOX” button. Chemicals can be searched by typing 3 letters of the chemical name which will then populate with the top 100 hits containing those 3 letters. The chemical will appear and display the CASRN number following the name. Up to 5 chemicals can be included in the ECOTOX filter. Unwanted chemicals can be individually removed by selecting chemical and then clicking the “Remove Selected Chemicals” or all chemicals can be removed by clicking the “Remove All Chemicals.” There are links to the “CompTox Chemical Dashboard” and “ECOTOX Chemicals,” which open the respective databases in separate browser tabs to aid in finding chemicals of interest. To push the created filtered group to the ECOTOX Explore page, click the “Open in ECOTOX” button. Clicking the button will open up a separate browser tab that will incorporate the user customized group within the ECOTOX webpage.

The selected species will be added to an ECOTOX Custom Species Group with the selected chemicals used as a filter in Explore for the user to view and download records from ECOTOX.

Chemical Search:  [CompTox Chemical Dashboard](#)

[ECOTOX Chemicals](#)

Selected Chemicals: Imidacloprid (CASRN:138261413)  
Flupyradifurone (CASRN:951659408)  
Thiachloprid (CASRN:111988499)

(3/5) CAS Numbers Selected

### Level 2: Functional Domain(s) Alignment

In the “View SeqAPASS Reports” tab, on the “Level 1 Query Protein Information” page, there is a “Level 2” box for comparing hit domains to the query domain. In the “Level 2” dropdown box, there is a link out to the “NCBI Conserved Domain Database” for the query protein of interest. Below this link the user will find a drop-down containing functional domains associated with the query sequence for comparison across species.

Main | Level 1 | DS Report

Level 1 Query Protein Information

Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list.  
SeqAPASS ID: 2419 Query Accession: [NP\\_001119.2](#) Ortholog Count: 712 Protein and Taxonomy Data: 04/28/2021  
Query Species: Homo sapiens BLAST Version: 2.13.0  
Query Protein: estrogen receptor isoform 1 Software Version: 0.1

Susceptibility Cut-off

Primary Report Settings

Visualization

Level 2

Level 2 Query Domain

[NCBI Conserved Domain Database](#)

Functional Domains

-Select Domain-

Choose Domain to View

-Select Completed Domain-

Level 3

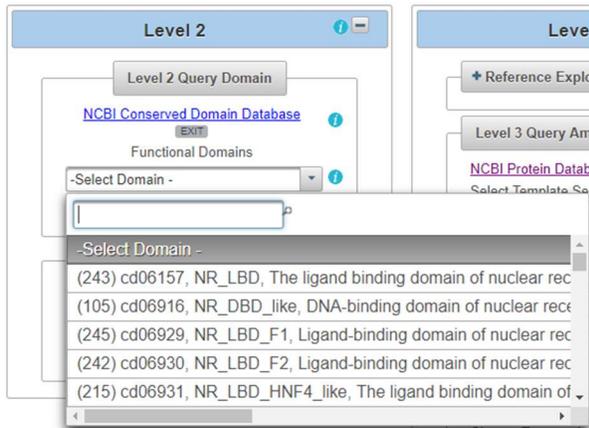
Level 4

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

In the drop-down box (below the words “Functional Domains”) the user will find all domains associated with the query protein listed in the “NCBI Conserved Domains Database”. To compare a domain from the query protein to domains of the hit proteins, the user will use the drop-down to highlight a domain and click the “Request Domain Run” button.

**Note:** Domains in the drop-down are listed with the first amino acid residue position that aligns with the NCBI curated domain in parenthesis, followed by the NCBI domain Accession, domain name, and description.



**Note:** The user can also use the text box on the top of the drop-down to search the “Functional Domain” list in the drop-down.

It is recommended that the user click on the “NCBI Conserved Domains Database”

<http://www.ncbi.nlm.nih.gov/cdd/> link to identify which domains are “Specific hits” in the NCBI Conserved Domains Database. On the NCBI page, the user can scroll over the graphical representation of the domains associated with the query sequence to highlight and identify the Accession associated with domain “Specific hits.” The example below shows the user hovering over the NR\_LBD\_ER domain with the computer mouse.

**Conserved domains on [gi|62821794|ref|NP\_000116.2]**

estrogen receptor isoform 1 [Homo sapiens]

**Graphical summary** Zoom to residue level show extra options >

**Query seq.** 1 100 200 300 400 500 595

**Specific hits** NR\_LBD\_ER

**Superfamilies** Oest\_recep NR\_LBD\_Like superfam

**List of domain hits**

Name	Accession	Description
NR_LBD_ER	cd06949	Ligand binding domain of Estrogen receptor which are activated by the hormone 17beta-estradiol (estrogen). The ligand-binding domain (LBD) of Estrogen receptor (ER), Estrogen receptor, a member of nuclear receptor superfamily, is activated by the hormone 17beta-estradiol (estrogen). Estrogen regulates many physiological processes including reproduction, bone integrity, cardiovascular health, and behavior. The main mechanism of action of the estrogen receptor element of target genes upon activation by estrogen and then recruiting coactivator proteins which are rapidly associated with other membrane proteins and can be rapidly activated by exposure of cells to estrogen ligand-activated transcription factors. ER has a central well conserved DNA binding domain (DBD), a variant binding domain (LBD). The C-terminal LBD also contains AF-2 activation motif, the dimerization motif, and part of the nuclear localization region. Estrogen receptor has been linked to aging, cancer, obesity and other diseases.

Pssm-ID: 132747 Cd Length: 235 Bit Score: 426.07 E-value: 1.45e-146

```

10 20 30 40 50 60 70 80
gi|62821794|310|LTADQMSVALLDAEPLVSEYDFRFFSEASMGLLTDLAARELVNINAKRVPFVQLTLRSDQVHLLCCAMLEIML 389
Cdd:cd06949 1 LSNELISALLAEAFPHIVSEYDFRFFSEASMGLLTDLAARELVNINAKRVPFVQLTLRSDQVHLLCCAMLEIML 80

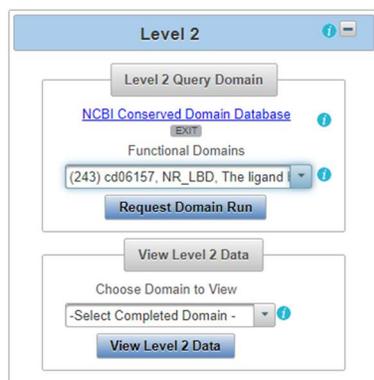
90 100 110 120 130 140 150 160
gi|62821794|390|GLVNRSMHFGKLLFAPNLLDRHQKVEYQVEITFMLLATSRFRMDLQGEFVCLAKSILLASGVTFLRSTKSL 469
Cdd:cd06949 81 GLVNRSMHFGKLLFAPNLLDRHQKVEYQVEITFMLLATSRFRMDLQGEFVCLAKSILLASGVTFLRSTKSL 157

170 180 190 200 210 220 230
gi|62821794|470|EKDQHIRVRLKIDITITLIMAKASLILQQQRSLQLLLLSHTRNRMNRMHGLYMGCRVYVLLLEMLDAR 547
  
```

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

After identifying the domain(s) of interest and the corresponding starting residue and domain Accession, the user can return to the SeqAPASS tool, scroll to the domain of interest in the drop-down. If that domain has not been previously run by the user, the “Request Domain Run” button will become active and the user can click it to submit the domain query.



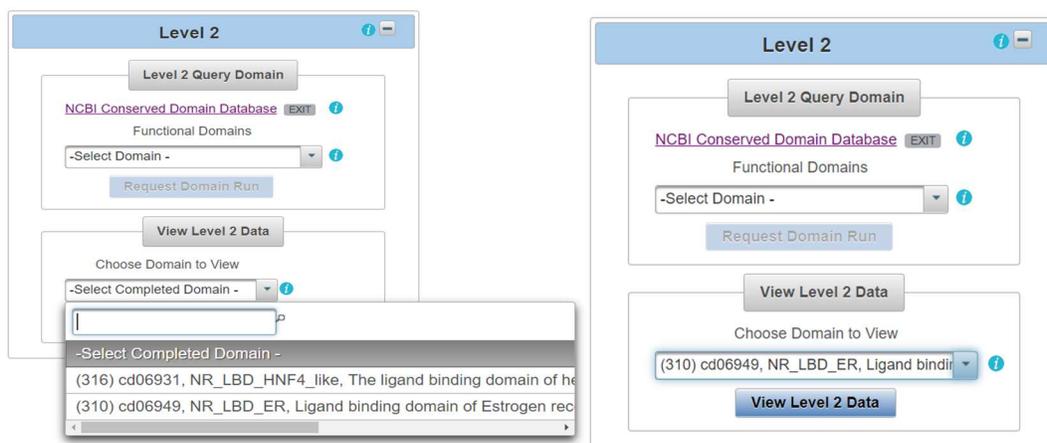
When user clicks the “Request Domain Run” button, the following message will appear if the runs has been submitted successfully.



When sequence comparisons have completed for the selected functional domain, the domain will be present in the “View Level 2 Data” drop-down. The *drop-down is not automatically populated* with the completed domain run. The *user must click on the “Refresh Level 2 and 3 runs” button to update the page* for the newly completed domain to present itself in the Choose Domain to View drop-down.

To view a completed Level 2 domain, highlight the domain of interest in the drop-down box and click the “View Level 2 Data” button. This will bring the user to the “Level 2” data page for the selected query protein/domain.

Note: The user can also use the text box on the top of the drop-down to search the “Completed Domain” list.

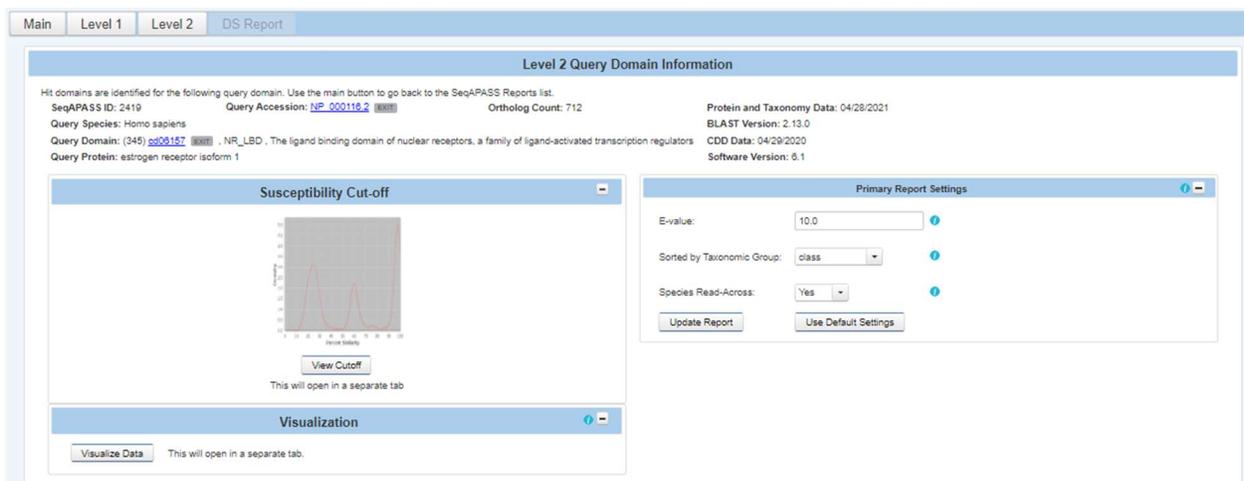


## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### View Level 2 Data Page

The “Level 2 Query Domain Information” box contains the SeqAPASS Run ID, Query Accession, Ortholog Count (# of hits identified as ortholog candidates to the query species protein sequence), NCBI Data updates (“Protein and Taxonomy Data:” and “CDD Data:” display the dates that NCBI databases were downloaded and incorporated into the SeqAPASS database; “BLAST version:” and “Software Version:” displays the version being used by the SeqAPASS tool for the selected data), Query Species, Query Domain (with link out to NCBI domain page), Query Protein name.



The default “Level 2” table is the “Primary Report”, which includes query domain information in the first row below the column titles, followed by hit domains whose sequences aligned with the selected query domain. The hit domains are ordered from the highest to lowest percent similarity (Maximum percent similarity =100%). For each hit domain, Data Version, NCBI Accession and species information is provided, including the “Protein Count” which indicates the number of protein records per species in the NCBI protein database, taxonomic information, and species names. Also included are the NCBI accession for the query protein, query protein name, Domain Type, BLASTp bitscore (describes overall quality of the alignment, See NCBI BLASTp tutorials), and Domain percent similarity ([hit bitscore/query bitscore]\*100). If the hit protein has been identified as an ortholog candidate (using reciprocal best hit BLAST method), it will be noted with a “Y” for yes or if not an ortholog candidate, a “N”, for no.

A prediction of susceptibility is displayed based on the susceptibility cut-off, identified with a “Y” for yes or an “N” for no. The date/time the analysis was completed is also identified. (See **Search, View, and Download Data Tables** section of user guide for more information). There is a column that identifies if the species is a eukaryote, noted with a “Y” for yes or alternatively a “N” for no if the hit is a prokaryote. Additionally, a column with a link to the U.S. EPA ECOTOX Knowledgebase (<https://cfpub.epa.gov/ecotox/help.cfm>) is available when there are empirical toxicity data curated for the species identified in the row. This link allows the user to view available single chemical toxicity data from the literature for specific species.

Default highlights identify partial protein sequences, sequences with a bitscore higher than the query domain and therefore percent similarity greater than 100% (commonly synthetic constructs), and when zero ortholog candidates are identified (in this case a user should consider a different query sequence). Additionally, the default setting for the report shows only eukaryote data, excluding prokaryote data from the table with the “Show Only Eukaryotes” checkbox checked. To view prokaryote data, deselect this checkbox.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Partial Hit Protein Sequence  
 Percent Similarity > 100%  
 Susceptible = 1: Estimated Count = 2  
 Show Only Eukaryotes

Primary Report  
 Full Report

[View Level 2 Summary Report](#)

---

**Level 2 Data - Primary**

The following links exit the site: [EXIT](#) [Download Current Level 2 Report Settings](#)

Search:

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name
4	<a href="#">NP_000116.2</a>	126506	<a href="#">9606</a>	Mammalia	Mammalia	<a href="#">Homo sapiens</a>	Human	<a href="#">estrogen receptor isoform 1</a>
4	<a href="#">ABY84717.1</a>	2023	<a href="#">9593</a>	Mammalia	Mammalia	<a href="#">Gorilla gorilla</a>	Western gorilla	<a href="#">estrogen receptor alpha</a>
4	<a href="#">XP_002817538.1</a>	145798	<a href="#">9601</a>	Mammalia	Mammalia	<a href="#">Pongo abelii</a>	Sumatran orangutan	<a href="#">estrogen receptor isoform X2</a>
4	<a href="#">XP_011852190.1</a>	38580	<a href="#">9588</a>	Mammalia	Mammalia	<a href="#">Mandrillus leucophaeus</a>	Drill	<a href="#">PREDICTED_estrogen receptor isoform X2</a>
4	<a href="#">XP_023061905.1</a>	54518	<a href="#">591936</a>	Mammalia	Mammalia	<a href="#">Ptilocolobus techoscèleus</a>	Ugandan red Colobus	<a href="#">estrogen receptor isoform X2</a>
4	<a href="#">XP_018884801.1</a>	47068	<a href="#">9595</a>	Mammalia	Mammalia	<a href="#">Gorilla gorilla gorilla</a>	Western lowland gorilla	<a href="#">PREDICTED_estrogen receptor isoform X2</a>
4	<a href="#">XP_008005788.1</a>	62315	<a href="#">60711</a>	Mammalia	Mammalia	<a href="#">Chlorocebus sabaeus</a>	Green monkey	<a href="#">PREDICTED_estrogen receptor isoform X2</a>
4	<a href="#">XP_011751932.1</a>	69122	<a href="#">9545</a>	Mammalia	Mammalia	<a href="#">Macaca nemestrina</a>	Pig-tailed macaque	<a href="#">estrogen receptor isoform X2</a>
4	<a href="#">ABY84719.1</a>	712	<a href="#">9580</a>	Mammalia	Mammalia	<a href="#">Hylobates lar</a>	Common gibbon	<a href="#">estrogen receptor alpha</a>
4	<a href="#">NP_001158059.1</a>	68224	<a href="#">9555</a>	Mammalia	Mammalia	<a href="#">Papio anubis</a>	Olive baboon	<a href="#">estrogen receptor</a>

(1 of 95) [Download Table:](#)

## Level Two Summary Report

The user can view a summary of the data for each taxonomic group by clicking on the “View Level 2 Summary Report”. The data includes, number of species, mean percent similarity, median percent similarity and susceptibility prediction. This data table can also be downloaded.

**Level Two Summary Report**

Taxonomic Group	Filtered Taxonomic Group	Number of Species	Mean Percent Similarity	Median Percent Similarity	Susceptibility Prediction
Mammalia	Mammalia	176	80.60	97.63	Y
Aves	Aves	96	83.78	95.73	Y
Crocodylia	Crocodylia	7	84.98	95.97	Y
Testudines	Testudines	9	86.30	94.55	Y
Lepidosauria	Lepidosauria	22	71.14	92.21	Y
Amphibia	Amphibia	22	60.74	81.03	Y
Chondrichthyes	Chondrichthyes	7	55.68	67.59	Y
Coelacanthiformes	Coelacanthiformes	2	70.43	70.43	Y
Actinopteri	Actinopteri	179	51.66	62.13	Y
Ceratodontimorpha	Ceratodontimorpha	3	53.96	71.15	Y

(1 of 6) [Download Table:](#)

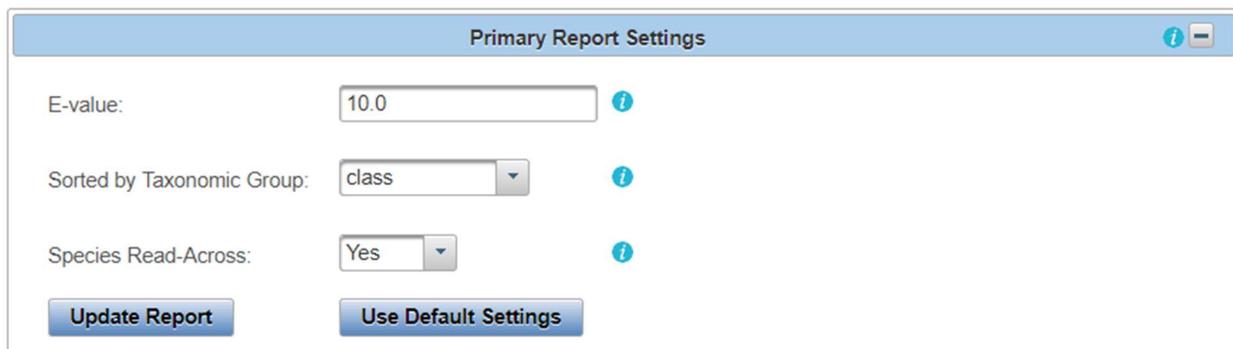
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Level 2: Primary Report Settings

#### Default settings

The “Primary Report Settings” box allows the user to view default settings on the table below and manipulate certain settings. The “Primary Report Settings” box is only available on the “Primary Report” display. The default settings show data for hits whose E-value are  $\leq 10$ . The default setting for the “Sorted by Taxonomic Group” is “class,” therefore the “Filtered Taxonomic Group” column in the table is set to identify and report the taxonomic lineage of “class” from the NCBI Taxonomy Database. However, if class is not identified in the NCBI Taxonomic Hierarchy associated with the hit accession, then the algorithm will report the next available Taxonomic Group moving from class to subclass, to superorder, to order, to suborder, to superfamily, to family, to subfamily, to genus. Finally, the susceptibility predictions are set by using Species Read-Across. (Please view **SeqAPASS Documentation** Section of the User Guide for details on Read-Across settings). Briefly, “Species Read-Across” is used to set the susceptibility prediction, where all ortholog candidates are Susceptible = Y; all species listed above the susceptibility cut-off are Susceptible = Y; all species below the cut-off from the same taxonomic group of one or more species above the cut-off are Susceptible = Y; and those below the cut-off that are not ortholog candidates and do not belong to a taxonomic group above the cut-off are Susceptible = N.



Primary Report Settings

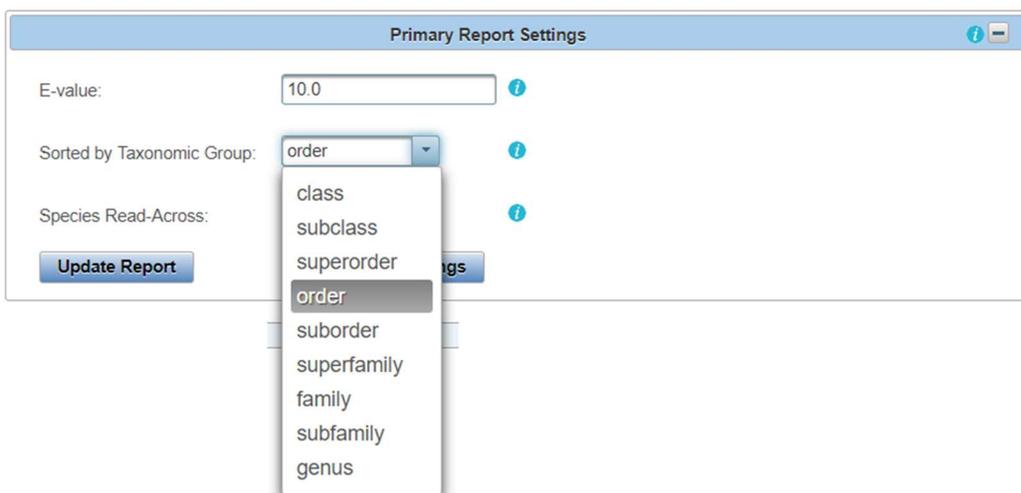
E-value:  ⓘ

Sorted by Taxonomic Group:  ⓘ

Species Read-Across:  ⓘ

#### Changing Default Settings

The user may choose to change the level of the taxonomic hierarchy that is used for the susceptibility prediction. From the “Sorted by Taxonomic Group” dropdown the user may choose to display a different taxonomic group in the “Filtered Taxonomic Group” column of the data table.



Primary Report Settings

E-value:  ⓘ

Sorted by Taxonomic Group:  ⓘ

Species Read-Across:  ⓘ

- class
- subclass
- superorder
- order**
- suborder
- superfamily
- family
- subfamily
- genus

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

If the user chooses “order” for example, the “Filtered Taxonomic Group” column in the data table will report the taxonomic lineage of “order” from the NCBI Taxonomy Database and all species read-across for the susceptibility prediction will be based on order instead of class. As described previously, if order is not identified in the NCBI Taxonomic Hierarchy associated with the hit accession, then the algorithm will report the next available Taxonomic Group moving from suborder, to superfamily, to family, to subfamily, to genus. Upon selecting the Taxonomic Group from the dropdown and clicking “Update Report,” the “Level 2” data for the Primary Report will update to the selected taxonomic level. The user can also download the currently applied report settings by selecting the “Download Current Level 2 Report Settings”. This csv file allows the user to track which settings were used or changed by the user when downloading a data table.

Level 2 Data - Primary							
The following links exit the site <a href="#">EXIT</a>							
Search: <input type="text" value="Enter keyword"/>							
Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name
4	<a href="#">NP_000116.2</a>	1265506	<a href="#">9806</a>	Mammalia	Primates	<a href="#">Homo sapiens</a>	Human
4	<a href="#">XP_014992596.1</a>	88400	<a href="#">9544</a>	Mammalia	Primates	<a href="#">Macaca mulatta</a>	Rhesus monkey
4	<a href="#">ABY64721.1</a>	931	<a href="#">9534</a>	Mammalia	Primates	<a href="#">Chlorocebus aethiops</a>	Grivet
4	<a href="#">XP_003255939.1</a>	38964	<a href="#">61853</a>	Mammalia	Primates	<a href="#">Nomascus leucogenys</a>	Northern white-cheeked gibbon
4	<a href="#">XP_025240309.1</a>	52618	<a href="#">9565</a>	Mammalia	Primates	<a href="#">Theropithecus gelada</a>	Gelada
4	<a href="#">XP_003811544.1</a>	51891	<a href="#">9597</a>	Mammalia	Primates	<a href="#">Pan paniscus</a>	Pygmy chimpanzee
4	<a href="#">XP_011922091.1</a>	66748	<a href="#">9531</a>	Mammalia	Primates	<a href="#">Cercocebus atys</a>	Sooty mangabey
4	<a href="#">ABY64717.1</a>	2023	<a href="#">9593</a>	Mammalia	Primates	<a href="#">Gorilla gorilla</a>	Western gorilla
4	<a href="#">XP_002817538.1</a>	145798	<a href="#">9801</a>	Mammalia	Primates	<a href="#">Pongo abelii</a>	Sumatran orangutan
4	<a href="#">XP_011852190.1</a>	38580	<a href="#">9568</a>	Mammalia	Primates	<a href="#">Mandrillus leucophaeus</a>	Drill

(1 of 95) 1 2 3 4 5 6 7 8 9 10 10 Download Table:

The user may also choose to turn species read across off, by using the “Species Read-Across” drop-down and selecting “No” and clicking “Update Report”. When “No” is selected, the susceptibility predictions will only be “Y” in the table below if Percent Similarity is above the Cut-off or if the hit is identified as an Ortholog Candidate, yes or “Y.” Any hit below the cut-off will yield a susceptibility prediction of no or “N”.

Primary Report Settings

E-value:

Sorted by Taxonomic Group:

Species Read-Across:

The user can select the “Full Report” on the “Level 2” data page, which includes the same information as the “Primary Report” and additional information pertaining to the alignment of the protein sequence using BLASTp and domain information. Additional information includes the NCBI PSSM ID, NCBI Domain ID, Domain Name, number of amino acid residues in the sequence (Hit Length), the number of exact matching amino acids between the hit and query sequence (Identity), the number of exact and similar (similar side-chain substitutions) matches in amino acids between the hit and the query sequence (Positives), and the expect value (E-value) describing the number of different alignments expected to



## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Susceptibility Cutoff Box for Level 2

The susceptibility prediction is set by identifying ortholog candidates, sequences above a defined susceptibility cutoff, or by identifying those species below the susceptibility cut-off from an organism class above the susceptibility cutoff. The default susceptibility cut-off is set by plotting the distribution of percent similarities calculated for each hit protein. From this plot, the critical points are identified, and the local minimums and maximums reported. Using the ortholog candidate data, a susceptibility cut-off is automatically determined by identifying the first ortholog candidate at an equal or higher percent similarity than the first local minimum percent similarity. The user can view this graph by clicking the “View Cutoff” button in the “Susceptibility Cut-off” box. Radio buttons located to the right of the graphical display indicate which Cut-off has been applied for the evaluation of susceptibility in the report. These radio buttons can be selected to change the cut-off in the table to the 2<sup>nd</sup> local minimum, where the 2<sup>nd</sup> local minimum is identified in the density plot and the first ortholog candidate at an equal or higher percent similarity than the second local minimum percent similarity is used to set the cut-off. Or the user can define the local minimum by clicking on the “User Defined” radio button. Alternatively, the user can view the closely examine the density plot and manipulate the cut-off by clicking the “View Cutoff” button.

#### Level 2 Query Domain Information

Hit domains are identified for the following query domain. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 1290	Query Accession: <a href="#">NP_000116.2</a> <a href="#">EXIT</a>	Ortholog Count: 348	Protein and Taxonomy Data: 02/28/2019
Query Species: Homo sapiens			BLAST Version: 2.8.1
Query Domain: (310) <a href="#">cd06949</a> <a href="#">EXIT</a> , NR_LBD_ER, Ligand binding domain of Estrogen receptor, which are activated by the hormone 17beta-estradiol (estrogen)			CDD Data: 12/08/2016
Query Protein: estrogen receptor isoform 1			Software Version: 3.2

#### Susceptibility Cut-off

[View Cutoff](#)

This will open in a separate tab

#### Visualization

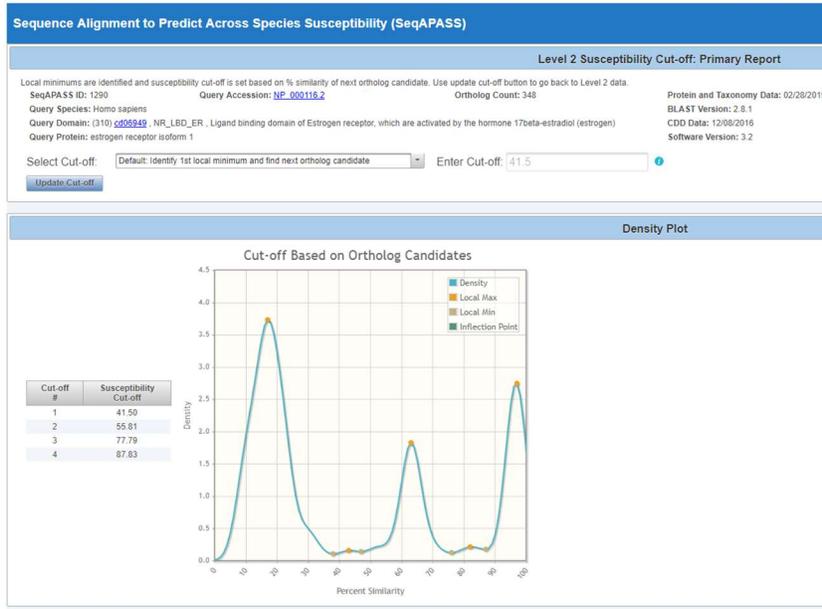
Upon clicking “View Cutoff” button, a new page is displayed with a drop-down that allows the user to set the susceptibility cut-off using the first local minimum and the identified ortholog candidate, the second local minimum and the identified ortholog candidate, or by the “User defined cut-off” (where the user selects the cutoff). To update the cut-off in the Level 2 data report and/or return to the Level 2 page, click “Update Cut-off” button.

**Note:** The user should have direct empirical evidence that species above the user defined cutoff are susceptible via the protein of interest, or that the species below the user defined cutoff are not susceptible.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Upon selecting the User defined cut-off from the dropdown, the “Enter Cut-off” text box becomes active and the user can enter a number 1-100.



All potential susceptibility cut-offs generated by the data distribution and ortholog candidate identification are reported in the table with columns “Cut-off #” and “Susceptibility Cut-off”. The user can use these numbers to define a cut-off if empirical evidence suggests that the “Default” or “2 minimum” are not supported.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## No Orthologs Detected

Level 2 Query Domain Information

Hit domains are identified for the following query domain. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 1326      Query Accession: [NP\\_001317544.1](#) (EXIT)      Ortholog Count: 0

Query Species: Homo sapiens

Query Domain: (110) [cd06965](#) (EXIT) . NR\_DBD\_Ppar . DNA-binding domain of peroxisome proliferator-activated receptors (PPAR) is composed of two C4-type zinc fingers

Query Protein: peroxisome proliferator-activated receptor gamma isoform 3

Susceptibility Cut-off

[View Cutoff](#)

This will open in a separate tab

Visualization i +

Partial Hit Protein Sequence i  
 Percent Similarity > 100%  
 Susceptible = Y, Ortholog Count = 0 [View Level 2 Summary Report](#)  
 Show Only Eukaryotes

Level 2 Data - Primary

The following links exit the site: [EXIT](#)

Search:  i

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name
4	<a href="#">NP_001317544.1</a>	1265506	9606	Mammalia	Mammalia	<a href="#">Homo sapiens</a>	Human
4	<a href="#">XP_008150376.1</a>	50340	29078	Mammalia	Mammalia	<a href="#">Eptesicus fuscus</a>	Big brown bat
4	<a href="#">XP_019283665.1</a>	58782	9691	Mammalia	Mammalia	<a href="#">Panthera pardus</a>	Leopard
4	<a href="#">XP_021047523.1</a>	36287	10093	Mammalia	Mammalia	<a href="#">Mus pabari</a>	Shrew mouse

If no orthologs are detected from reciprocal best hit blast analysis, the “Ortholog Count” will be “0” at the top of the “Level 2 Query Protein Information” page. The cutoff will be set by the local minimums only, therefore the susceptibility prediction will NOT take into account ortholog candidates. ***It is recommended that the user checks the full report for Ortholog candidates or identifies a different query sequence for the susceptibility predictions.*** Here, the susceptibility predictions will be highlighted in dark pink in the Level 2 data table to indicate that 0 orthologs were detected and the susceptibility cutoff was determined from plotting the distribution of percent similarities and identifying the local minimums.

Main Level 1 Level 2

Level 2 Query Domain Information

Hit domains are identified for the following query domain. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 1321      Query Accession: [BAF57671.1](#) (EXIT)      Ortholog Count: 0      Protein and Taxonomy Data: 02/28/2019

Query Species: Mus caroli      BLAST Version: 2.8.1

Query Domain: (24) [CHL00070](#) (EXIT) . pe1B . cytochrome b6      CDD Data: 12/08/2016

Query Protein: cytochrome b, partial (mitochondrion)      Software Version: 4.0

Susceptibility Cut-off +

Primary Report Settings i +

Visualization i +

By clicking on the “View Cutoff” button when no orthologs are detected, the “Cut-off #” and “Susceptibility Cut-off” columns will report only the local minimum values.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

**Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)**

**Level 2 Susceptibility Cut-off: Primary Report**

Local minimums are identified and susceptibility cut-off is set based on % similarity of next ortholog candidate. Use update cut-off button to go back to Level 2 data.

SeqAPASS ID: 1326      Query Accession: [NP\\_001317544.1](#)      Ortholog Count: 0

Query Species: Homo sapiens      Protein and Taxonomy Data: 02/28/2019  
 BLAST Version: 2.8.1  
 CDD Data: 12/08/2016  
 Software Version: 4.0

Query Domain: (110) [c00696S](#), NR\_DBD\_Ppar, DNA-binding domain of peroxisome proliferator-activated receptors (PPAR) is composed of two C4-type zinc fingers.  
 Query Protein: peroxisome proliferator-activated receptor gamma isoform 3

Select Cut-off:  Enter Cut-off:

[Update Cut-off](#)

---

**Density Plot**

Cut-off Based on Ortholog Candidates

The user can return to the “Level 2” data page by clicking the “Update Cut-off” button or exiting the tab.

## Level 1 and Level 2: Data Visualization

From the Level 1 or Level 2-results page SeqAPASS users can access an interactive data visualization for both the “Primary Report” or “Full Report” by clicking on the “Visualize Data” button.

Example of Level 1 page:

Home Request SeqAPASS Run SeqAPASS Run Status View SeqAPASS Reports Settings

SeqAPASS Reports Version 6.0 Logged in as: Donovan Blatz

Main Level 1 DS Report

**Level 1 Query Protein Information**

Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 2295      Query Accession: [NP\\_000116.2](#)      Ortholog Count: 656      Protein and Taxonomy Data: 04/28/2021  
 Query Species: Homo sapiens      BLAST Version: 2.11.0  
 Query Protein: estrogen receptor isoform 1      Software Version: 5.1

**Susceptibility Cut-off**

[Cutoff Settings](#)

This will open in a separate tab

Level 2

[Refresh Level 2 and 3 runs](#)

Level 3

**Primary Report Settings**

E-value:

Sorted by Taxonomic Group:

Common Domains:

Species Read-Across:

[Update Report](#)      [Use Default Settings](#)

**Visualization**

[Visualize Data](#)      This will open in a separate tab

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Example of Level 2 page:

The screenshot displays the 'Level 2 Query Domain Information' page. At the top, there are navigation tabs for 'Main', 'Level 1', 'Level 2', and 'DS Report'. The main content area is titled 'Level 2 Query Domain Information' and contains the following information:

- Hit domains are identified for the following query domain. Use the main button to go back to the SeqAPASS Reports list.
- SeqAPASS ID: 2419
- Query Accession: [NP\\_000116.2](#)
- Ortholog Count: 712
- Protein and Taxonomy Data: 04/28/2021
- Query Species: Homo sapiens
- BLAST Version: 2.13.0
- Query Domain: (345) [cd00167](#), NR\_LBD, The ligand binding domain of nuclear receptors, a family of ligand-activated transcription regulators
- CDD Data: 04/29/2020
- Query Protein: estrogen receptor isoform 1
- Software Version: 6.1

Below this information are two panels:

- Susceptibility Cut-off:** Contains a line graph showing percent similarity across species. A 'View Cutoff' button is present, with a note: 'This will open in a separate tab.'
- Primary Report Settings:** Contains input fields for 'E-value' (set to 10.0), 'Sorted by Taxonomic Group' (set to 'class'), and 'Species Read-Across' (set to 'Yes'). It includes 'Update Report' and 'Use Default Settings' buttons.

At the bottom, there is a 'Visualization' panel with a 'Visualize Data' button and a note: 'This will open in a separate tab.'

The data visualization will then open in a new web browser tab, one for Level 1 and a different one for Level 2. The visualization will display for the report selected by the user on the Level 1 or Level 2 report page and be identified as “Level One Visualization – Primary Report” or “Level One Visualization – Full Report” and “Level Two Visualization – Primary Report” or “Level Two Visualization – Full Report.”

**Note:** One report type at a time, either “Primary Report” or “Full Report,” can be displayed in the visualization tab for Level 1 and Level 2. Therefore, if the user is viewing the “Level One Visualization – Primary Report” page and returns to the Level 1 results page and clicks the radio button for “Full Report,” the data visualization tab will update to “Level One Visualization – Full Report.”

## Level 1 and 2 Information Page

The initial page that opens upon clicking the “Visualize Data” button provides the respective level query protein information, including SeqAPASS ID, query protein, query species, ortholog count, and query accession information. A link out to the NCBI protein database page corresponding to the queried accession is available by clicking the query accession. Information on the visualization is provided in the “Visualization Info” text box. To view the data visualization boxplots click the BoxPlot icon.

The screenshot displays the 'Level One Visualization - Primary Report' page. At the top, there are navigation tabs for 'Main', 'Level 1', 'Level 2', and 'DS Report'. The main content area is titled 'Level 1 Query Protein Information' and contains the following information:

- SeqAPASS ID: 120
- Query Accession: [IC\\_X111.2](#)
- Query Protein: estrogen receptor isoform 1
- Query Species: Homo sapiens
- Ortholog Count: 310

Below this information is a section titled 'Select to Open Information or Data Visualization' with a 'BoxPlot' icon. Below that is an 'Info' section with a 'Visualization info' text box containing the following information:

The following data visualization is available for Level 1 and Level 2 data:

- BoxPlot** - Boxplots depicting SeqAPASS data illustrating the percent similarity across species compared to the query species examining the primary amino acid sequences (Level 1 Visualization) or functional domain (Level 2 Visualization).
  - The open circle, o, represents the query species and closed circles, •, represent the species with the highest percent similarity within the specified taxonomic group.
  - The top and bottom of each box represent the 75th and 25th percentiles, respectively. The top and bottom whiskers extend to 1.5 times the interquartile range.
  - The mean and median values for each taxonomic group are represented by horizontal thick and thin black lines on the box, respectively.
  - The dashed line indicates the cut-off for susceptibility predictions (based on ortholog analysis).



# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Level 1 and 2 BoxPlot Page – Controls

Upon clicking the “BoxPlot” icon on either Level 1 or Level 2 Visualization Information pages, a box for the boxplot “Controls” and a box for the interactive boxplot will open, respectively.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

**Level Two Visualization - Primary Report**

Level 1 Query Protein Information

SeqAPASS ID: 1290      Query Accession: [NP\\_000116.2](#)  
 Query Species: Homo sapiens      Ortholog Count: 348  
 Query Domain: (310) [cd06949](#), NR\_LBD\_ER, Ligand binding domain of Estrogen receptor, which are activated by the hormone 17beta-estradiol (estrogen)

Select to Open Information or Data Visualization

BoxPlot

Controls

**Taxonomic Groups (x-axis labels)**

Mammalia	Crocodylia	Aves	Testudines	Lepidosauria	Amphibia	Chondrichthyes	Coelacanthiformes	Actinopteri	Ceratodontiforma	
Cladistia	Myxiniiformes	Petromyzontiformes	Bivalvia	Branchiostomidae	Gastropoda	Enteropneusta	Priapulimorpha	Ascidiacea		
Cephalopoda	Polychaeta	Arachnida	Malacostraca	Insecta	Colembola	Hexanauplia	Litopsida	Plidiophora	Lingulata	Enopla
Ciliellata	Echinoidea	Branchiopoda	Holothuroidea	Eutardigrada	Merostomata	Asteroidea	Anthozoa	Polyplacophora	Chilopoda	
Monogononta	Rhopaluriidae	Hydrozoa	Trichoplacidae	Scyphozoa	Cubozoa	Chromadorea	Udeonychophora	Rhabditophora	Trematoda	
Enopla	Appendicularia	Cestoda	Diplopoda							

**Select Species for Legend:**

**Species Legend Options:**  
 Common Name     Group by Common Name  
 Scientific Name

**Optional Selections:**  
 Ortholog Candidates     Threatened Species     Endangered Species     Common Model Organisms

Boxplot

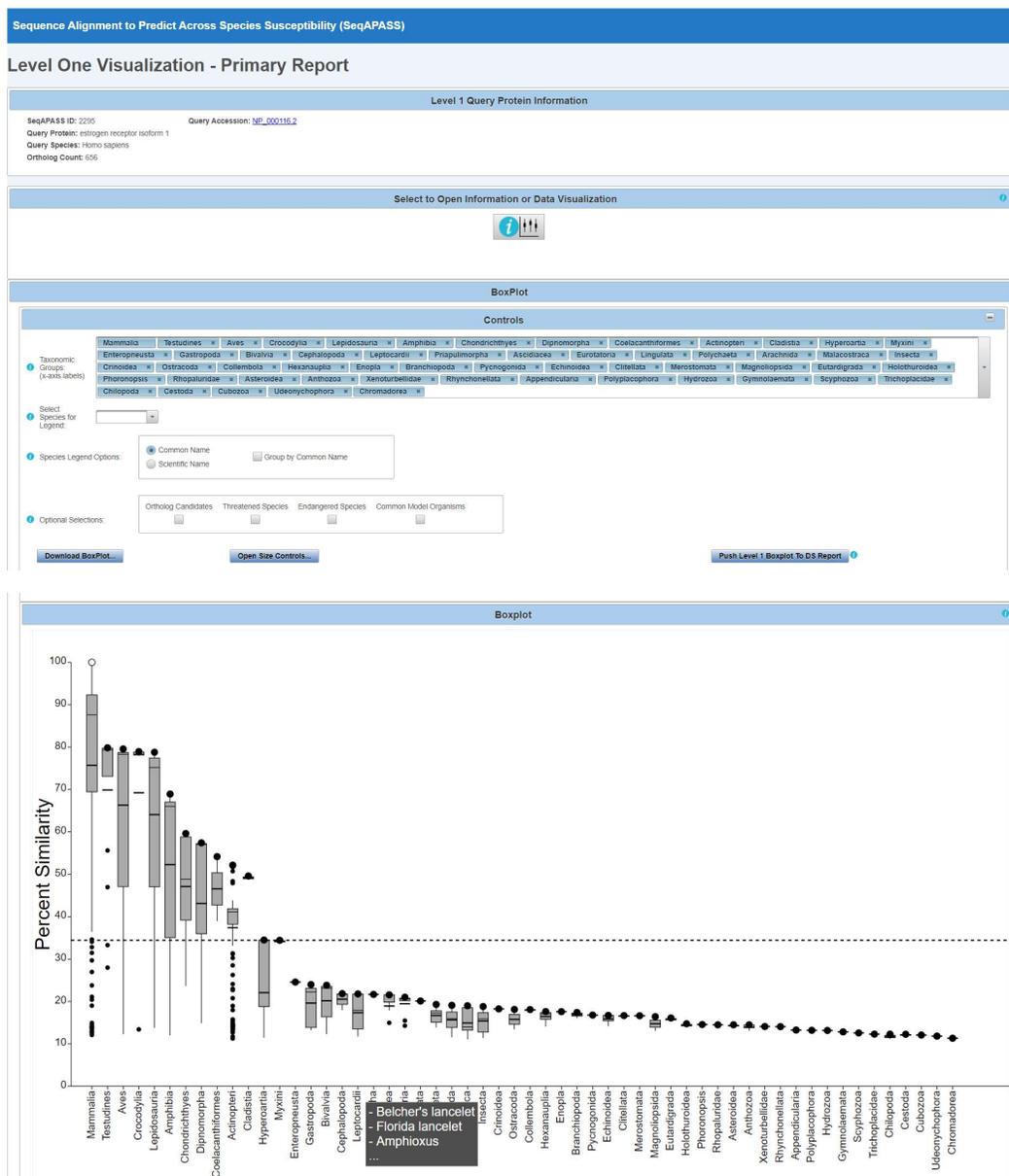
**Taxon**

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Manipulating Taxonomic Groups on x-axis

The boxplot controls allow the user to edit the taxonomic groups that are displayed on the x-axis by clicking on the “X” for the Taxonomic Group name (e.g., Aves). This action removes the selected group from the x-axis. To the right of the “Taxonomic Groups” controls box is a drop-down that allows the user to remove or add back taxonomic groups to the x-axis of the boxplot graphic, by deselecting or selecting check-boxes in the dropdown. Similarly, unwanted taxonomic groups may be removed directly from the boxplot by hovering the cursor over the taxonomic groups listed along the x-axis. The user will notice that the selection arrow changes to a black arrow with a red ‘x’ next to it; clicking the taxonomic group will then remove it from the boxplot and the “Taxonomic Groups” controls box. The user can delete multiple species by pressing CTRL and either clicking individual species or slowly dragging across multiple species. Additionally, that taxonomic group will have the checkbox deselected in the “Taxonomic Groups” controls box drop-down list.



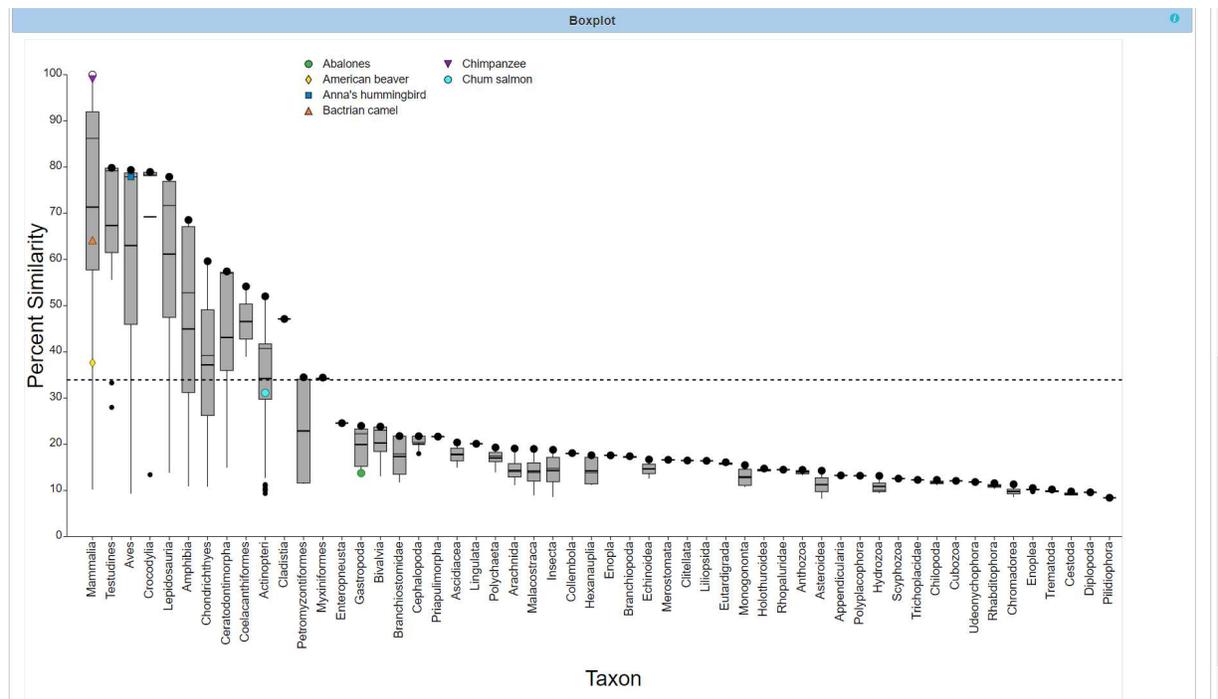
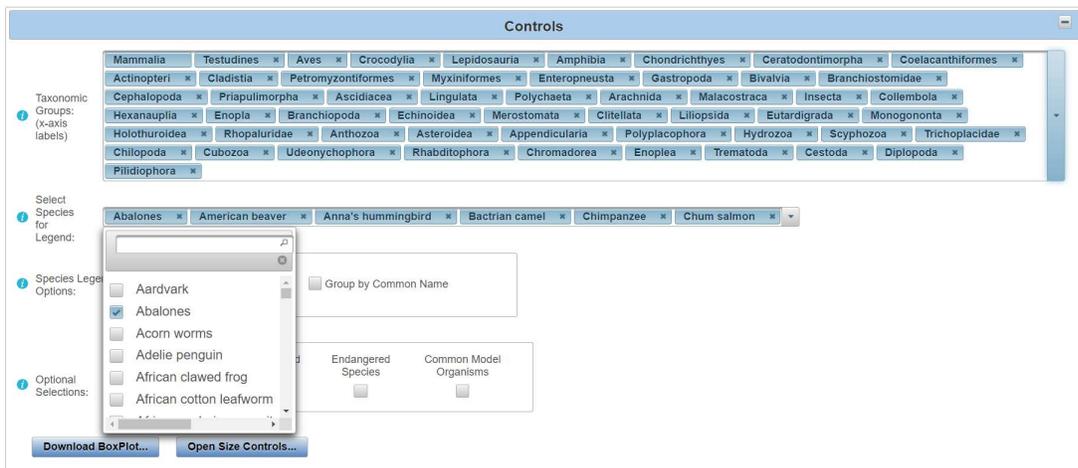
# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Customize Boxplot Legend

The user may customize the “Boxplot” by adding a legend that will pinpoint species of interest on the boxplot. Upon clicking the drop-down for “Select Species for Legend” in the controls box the user may search in the text box for specific species to display in the boxplot legend. Upon identifying a species from the drop-down menu and selecting the checkbox the species name will be placed in the boxplot legend and a corresponding data point will be produced on the graph. The default settings display the species common name both in the “Select Species for Legend” dropdown and on the boxplot. However, if the species scientific name is desired, the user can select the radio button for “Scientific Name” in the controls box for “Species Legend Options.” This action will change the drop-down menu and species in the legend to display the species scientific name.

**Note:** The database will take a brief moment to update the list upon changing between “Common Name” and “Scientific Name.”



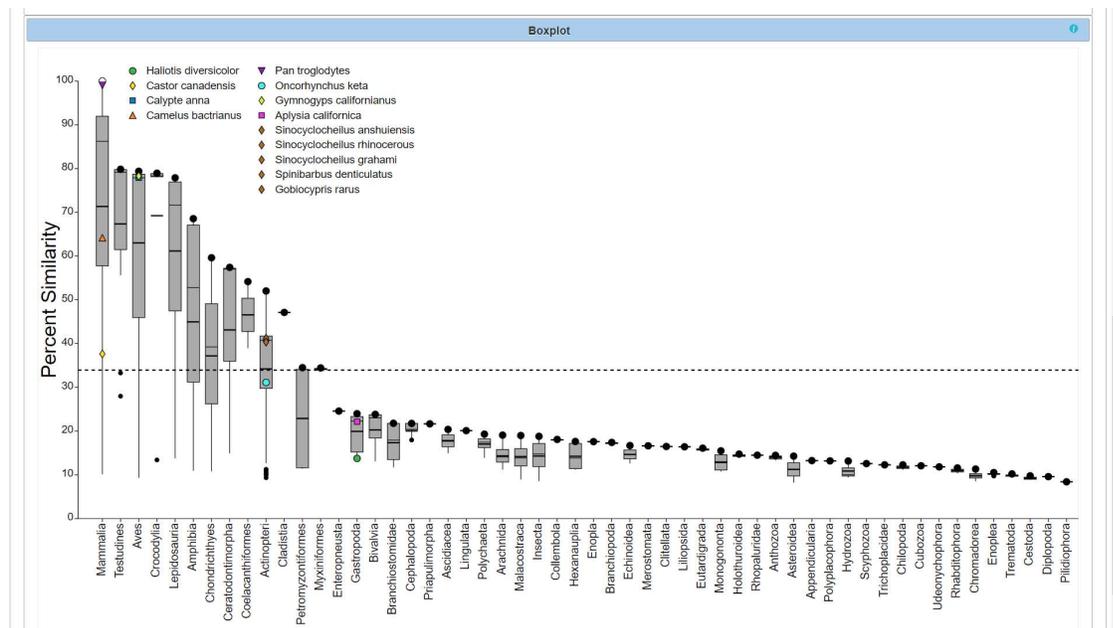
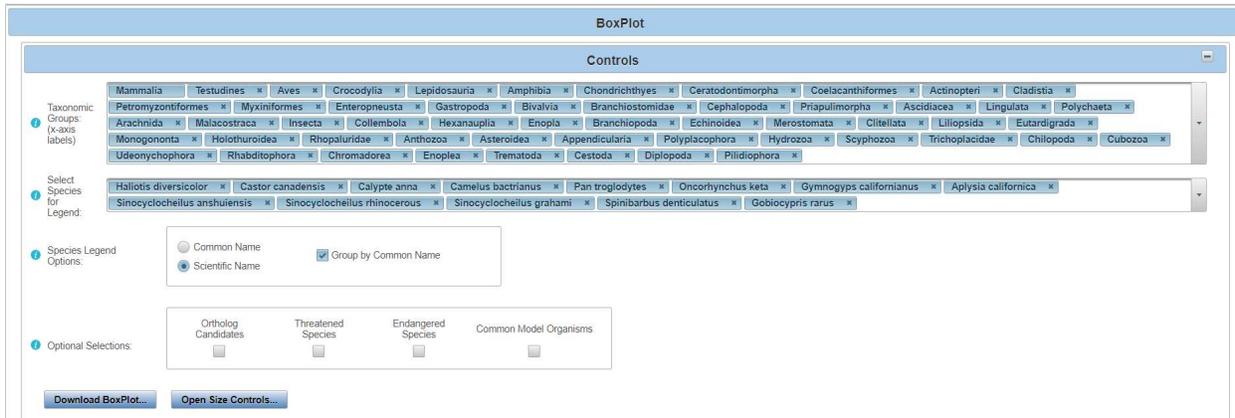
# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Change Species Display on Plot

Multiple scientific names can be represented by only one common name (e.g., Common name: Teleost fishes; corresponding scientific names: *Spinibarbus denticulatus*, *Sinocyclocheilus rhinoceros*, *Sinocyclocheilus grahami*, *Sinocyclocheilus anshuiensis*, *Gobiocypris rarus*, *Thamnaconus septentrionalis*). Therefore, if a species common name that represents multiple species was used to create the legend, and the user decides to instead select “Scientific Name,” by default the boxplot legend will change to display multiple scientific names that representing the individual common name and each scientific name will be represented by a unique color/shape point on the plot. However, if the user selects the checkbox “Group by Common Name” in the “Species Legend Options” control box, then the scientific names that are represented by one common name will all display the same color/shape point on the plot.

The user has the option of removing selected species from the legend either by removing them directly from the “Select Species for Legend” drop-down box or by hovering the mouse directly over the species name in the legend. The mouse will change to a black arrow with a red ‘x’ next to it. Clicking the name while this arrow is displayed will remove the species from the legend and from the control box.



# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

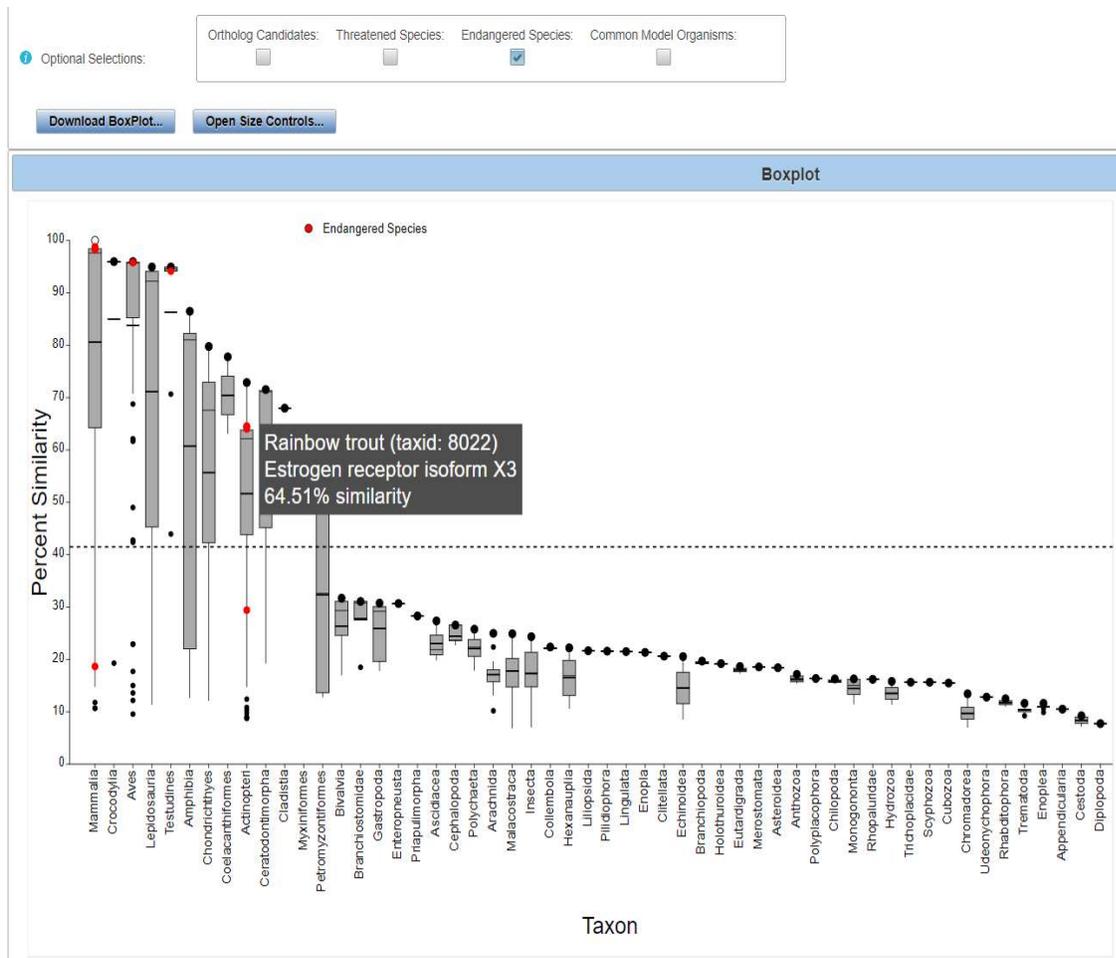
Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Customize the Legend to Display Species Groups of Interest

In the “Optional Selections” controls box, the user has the option of displaying “Ortholog Candidates,” “Threatened Species,” “Endangered Species,” or “Common Model Organisms.” Upon selecting one of the checkboxes, red data points corresponding to species will be displayed on the boxplot. By hovering the mouse over a single red point, a pop-up box will appear with the corresponding species name, taxonomic ID, query protein, and percent similarity.

**Note:** The user can select to display either species common name or scientific name in the hover over information box by selecting from the “Species Legend Options.”

If the user selects either “Threatened Species” or “Endangered Species,” clicking on an individual red dot will open a new web browser tab and link to the corresponding species page on the US Fish and Wildlife Service’s Environmental Conservation Online System (USFWS, ECOS; e.g.,) (<https://ecos.fws.gov/ecp0/profile/speciesProfile?sId=1506>).



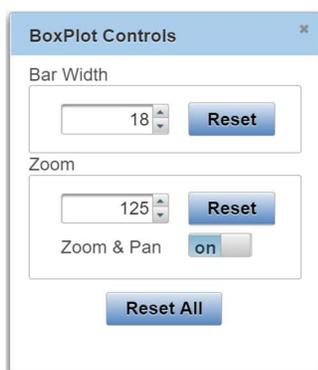
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### *BoxPlot Controls Widget for Bar Width, Zoom and Pan*

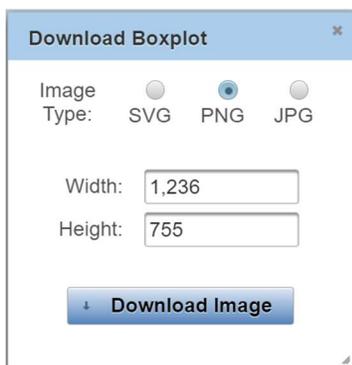
By clicking the “Open Size Controls” button, a “BoxPlot Controls” widget opens that allows the user to adjust the size of the bars on the boxplot by increasing or decreasing the “Bar Width” using the up and down arrows. The minimum and maximum size for bars are 6 and 60, respectively. To reset the bar width on the boxplot to default size, click the “Reset” button to the right of the “Bar Width” adjustment box in the “BoxPlot Controls” box. The user can also Zoom and Pan the boxplot by toggling the on /off button under the “Zoom” heading. The user can then zoom in or out by clicking the up or down arrows or entering a number in the text box and clicking enter. To reset the zoom on the boxplot to default size, click the “Reset” button to the right of the “Zoom” adjustment box in the “BoxPlot Controls” widget. The pan option is available when the “Zoom and Pan” option is toggled to the “on” position, which allows the user to click on the boxplot and drag the plot around the screen to reposition. To reset all BoxPlot Controls to default settings click the “Reset All” button.

**Note:** Upon exiting out of the BoxPlot Controls widget, the Zoom and Pan options are automatically turned off.



### *Download BoxPlot Widget*

To download the boxplot, click “Download BoxPlot” button in the controls box. A “Download Boxplot” Widget will pop up. It will be necessary to specify which type of file (SVG, PNG, or JPG,) to downloaded by clicking on the desired radio button for “Image Type.” The user may customize the resolution of the boxplot for PNG and JPG files prior to download by altering the “Width” and “Height” of the BoxPlot. To change “Width” or “Height,” enter the desired number in the text boxes. Click “Download Image” button to download the file. To close the “Download Boxplot” widget, click the “x” on the top right of the widget.



## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### *Hover-over Features in the BoxPlot*

By hovering over a taxonomic group name on the x-axis of the boxplot, an information box will pop-up listing the top three species in order by highest percent similarity. If only one or two species are represented in the taxonomic group, then only those species will be displayed. Hovering the mouse over any of the species in the boxplot, that is present in the legend, will generate a pop-up box with the corresponding species name, taxonomic ID, query protein, and percent similarity. The susceptibility cut-off is displayed in a pop-up text box upon hovering over the dashed horizontal cut-off line.

### *Summary Table for Species in a Specific Taxonomic Group*

By clicking on a box representing a taxonomic group in the boxplot a table will pop-up providing summary information for that particular group. The table header will provide summary statistics (i.e., mean and median percent similarity), including the Taxonomic Group name, number of species represented in the box, the overall susceptibility prediction for the selected taxonomic group. Data table includes protein and species information along with metrics for evaluated protein similarity and predicting susceptibility. Also included in the table are columns indicating if a species belongs to a certain group of interest (e.g., Threatened Species; Endangered Species, Model Organism). Table can be downloaded by clicking on the icon for excel or csv file.

### *Interactive Visualization with Level 1 Data Page and Level 2 Data Page*

The data visualization is programmed to update with changes made to the Level 1 Data page and Level 1 Data page, respectively. Therefore, if the user updates the Susceptibility Cut-off (See user guide section **Susceptibility Cutoff Box for Level 1** and **Susceptibility Cutoff Box for Level 2**) to the “Second Local Minimum” or “User Defined Cut-off,” the previously opened data visualization boxplot tab will update the cut-off accordingly. Similarly, the user modifies the Primary Report Settings (See user guide section **Level 1: Primary Report Settings** and **Level 2: Primary Report Settings**), the data visualization will update accordingly.

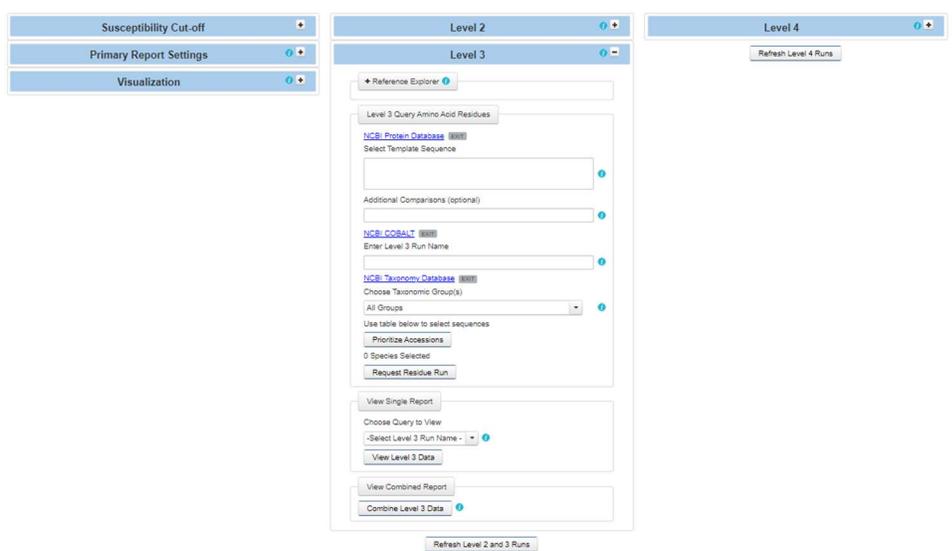
Note: If the user updates the “Primary Report Settings” for “Sorted by Taxonomic Group” the boxplot will update to display the new taxonomic group selection that is present in the “Filtered Taxonomic Group” column in the data table. The user should be aware that manipulating the “Sorted by Taxonomic Group” to a different level in the taxonomic lineage (e.g., from class to order; from class to genus) adds a larger number of taxonomic groups to the x-axis. Therefore, the plot may require greater user manipulation using the “BoxPlot Controls” to view the data.

### **Level 3: Individual Amino Acid Residue Alignment**

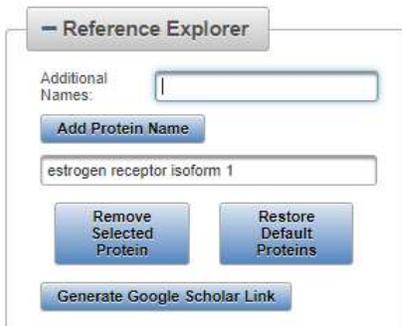
In the “View SeqAPASS Reports” tab, on the “Level 1 Query Protein Information” page, there is a “Level 3” dropdown for setting up the query for comparing individual amino acid residues to a template sequence. It is anticipated that the choice of template sequence and residues that are selected to align will be derived from the published literature in most cases. Publications evaluating homology models, protein crystal structures, pesticide field resistance, or utilizing site-directed mutagenesis are a few examples of the types of studies that may contain such information to guide a Level 3 SeqAPASS evaluation.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

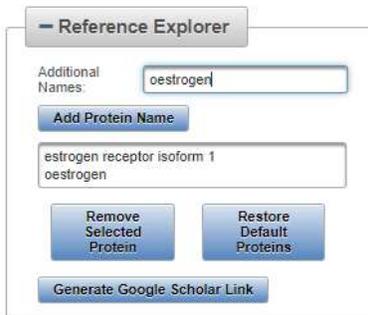
Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov



Relevant literature containing these data can be identified using the SeqAPASS “Reference Explorer.” The user can search for literature with the protein(s) of interest with an auto-populated search term that is integrated into a predefined Boolean string and generate a Google Scholar link that will take them to scientific articles containing their protein(s).



The user can modify the Boolean search string by adding text to the “Additional Names” text box and clicking the “Add Protein Name” button. By selecting a name that is currently in the text box and clicking the “Remove Selected Protein” button, the user can delete names from the text box and therefore these names will not be included in the Boolean string for the Google Scholar search.

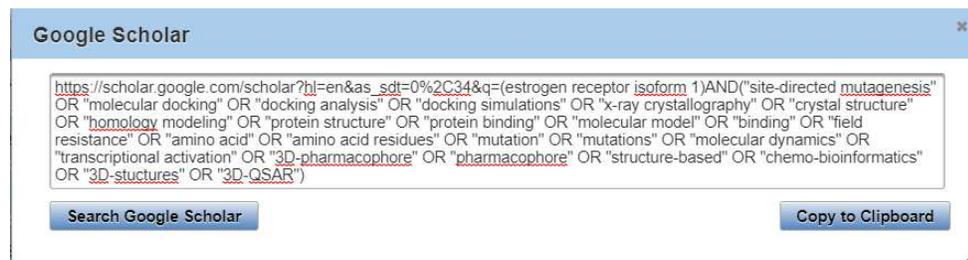


When satisfied with the protein names to be included in the Boolean search string, the user will select the “Generate Google Scholar Link” button. A pop-up will appear displaying the Boolean string to be searched in Google Scholar. The user can continue to modify the Boolean string by clicking in the text and adding additional information. The Boolean string can be copied and pasted elsewhere by the user by

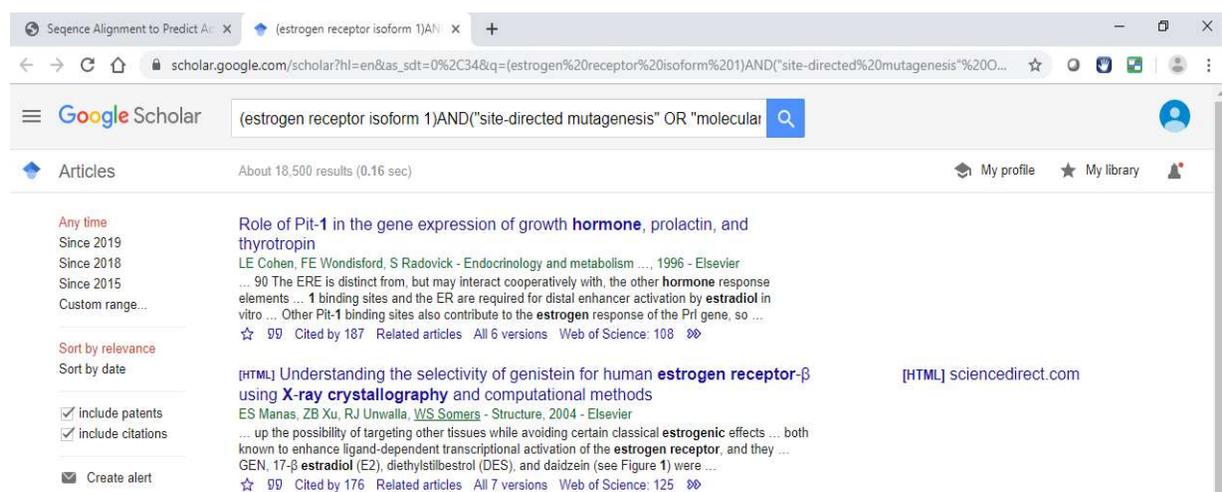
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

clicking the “Copy to Clipboard” button. The user can also choose to use the generated Boolean string to search Google Scholar. To do so the user will select the “Search Google Scholar” button.



Upon selecting the “Search Google Scholar” button, a new tab will be generated in the browser for Google Scholar that contains the Boolean string in the search with publications and articles that matched the SeqAPASS generated Boolean string. The literature displayed by Google Scholar for the user should be evaluated to identify appropriate articles for determining Level 3 template sequences and critical individual amino acids for comparisons across species.



In the “Level 3” box, there is a link out to the “NCBI Protein Database” for identifying the template sequence of interest. Below this link the user will find a text box where the user can enter an NCBI Protein Accession with the version number (e.g., NP\_000116.2) or a FASTA formatted sequence (e.g., <pre>gi|62821794|ref|NP\_000116.2| estrogen receptor isoform 1 [Homo sapiens]
MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPAVYNYPEGAAAYEFNA
AAAANA
QVYGQTGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLENE
PSGYT
VREAGPPAFYRPNSDNRQGGRERLASTNDKGSMMAMESAKETRYCAVCNDYASGYHYGVWSC
EGCKAFFK
RSIQGHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRD
DGEGRGEV
GSAGDMRAANLWPSPLMIKRSKKNLALSALTADQMVSALLDAEPPILYSEYDPTRPFSEASMMG
LLTNLA
DRELVHMINWAKRVPGFVDLTLHDQV).</pre>

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Upon clicking on in the “Select Template Sequence” text box, a pop-up message will appear to provide examples for the proper format of Accessions or FASTA files to be entered. A link out to the NCBI Protein Database is available for the user and found above the template entry text box.

Level 3

+ Reference Explorer ⓘ

Level 3 Query Amino Acid Residues

[NCBI Protein Database](#) ⓘ

Select Template Sequence

Additional Comparisons (optional)

[NCBI COBALT](#) ⓘ

Enter Level 3 Run Name

[NCBI Taxonomy Database](#) ⓘ

Choose Taxonomic Group(s)

All Groups

Use table below to select sequences

Prioritize Accessions

0 Species Selected

Request Residue Run

View Single Report

Choose Query to View

-Select Level 3 Run Name - ⓘ

View Level 3 Data

View Combined Report

Combine Level 3 Data ⓘ

Refresh Level 2 and 3 Runs

-Enter NCBI Protein Accession OR FASTA Sequence-  
Examples:  
NP\_000116.2  
OR  
>Sequence description in first line  
MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGVEVYLDSSKPAVY

Additional sequences can (this is an optional field the user can choose to fill in) also be incorporated into the Level 3 alignment using the “Additional Comparisons (optional)” text box. Upon clicking on the “Additional Comparisons (optional)” text box, a pop-up message will appear to provide examples for the proper format of Accessions or FASTA files to be entered.

Note: In the “Additional Comparisons (optional)” text box, zero or more NCBI Protein Accession must be entered prior to FASTA sequence(s) if they are to be included in the Level 3 alignment.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

The screenshot displays the SeqAPASS web interface. On the left, the 'Level 3' panel is active, showing options for 'Reference Explorer', 'Level 3 Query Amino Acid Residues', 'NCBI Protein Database', 'Select Template Sequence', 'Additional Comparisons (optional)', 'NCBI COBALT', 'Enter Level 3 Run Name', 'NCBI Taxonomy Database', 'Choose Taxonomic Group(s)', 'All Groups', 'Use table below to select sequences', 'Prioritize Accessions', '0 Species Selected', 'Request Residue Run', 'View Single Report', 'Choose Query to View', '-Select Level 3 Run Name -', 'View Level 3 Data', 'View Combined Report', and 'Combine Level 3 Data'. A 'Refresh Level 2 and 3 Runs' button is at the bottom. On the right, the 'Level 4' panel is visible with a 'Refresh Level 4 Runs' button. A tooltip over the 'Additional Comparisons (optional)' field contains the following text:

```
-Enter 0 or more NCBI Protein Accession(s) followed by 0 or more FASTA Sequence(s)-  
Examples:  
NP_000116.2  
1JLY_A  
>Sequence description of first FASTA  
MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPAVY  
>Sequence description of second FASTA  
XAGLPVIMCLKSNHQQYLRYQSDNIQQYGLLQFSADKILDPLAQFEVEPSKTYDGLV
```

Below the text box where the user can choose to add additional sequences for comparison, is a link to NCBI COBALT (Constraint-based Multiple Protein Alignment Tool). The NCBI COBALT allows the user to align multiple sequences and is the alignment tool that SeqAPASS algorithms utilize to set up the query of individual amino acid residues across species.

Note: The user does not need to use the COBALT link to run a Level 3 evaluation, however the link is available in case the user chooses to further evaluate or compare multiple potential template sequences.

Under the text “Enter Level 3 Run Name,” there is a text box where the user can enter a user defined name for the run. The user may only enter letters or integers as text for the name. The user defined name will appear in the “View Level 3 Data” dropdown upon completion of the Level 3 sequence alignment.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Level 3

+ Reference Explorer

Level 3 Query Amino Acid Residues

[NCBI Protein Database](#) EXIT

Select Template Sequence

Additional Comparisons (optional)

[NCBI COBALT](#) EXIT

Enter Level 3 Run Name

[NCBI Taxonomy Database](#) EXIT

Choose Taxonomic Group(s)

All Groups

Use table below to select sequences

Prioritize Accessions

0 Species Selected

Request Residue Run

View Single Report

Choose Query to View

-Select Level 3 Run Name -

View Level 3 Data

View Combined Report

Combine Level 3 Data

Refresh Level 2 and 3 Runs

To complete the set-up for a Level 3 query the user must select which sequences to compare to the identified template sequence. Listed in the “Choose Taxonomic Group(s)” drop-down are all Taxonomic Groups that were identified as hits in the “Level 1” primary amino acid sequence alignment data. Because COBALT is used to align all sequences that are selected, it is recommended that the user selectively identify sequences from the hit table below to align. For example, selecting sequences with low similarity to the template sequence along with sequences sharing high similarity to the template sequence can skew the alignment because COBALT is trying to align all the sequences together. It is recommended that the user select sequences by first selecting a taxonomic group from the “Choose Taxonomic Group(s)” drop-down. The user can also use the NCBI taxonomy link to type in the name of the “Taxonomic Groups” found in the drop-down to look up which species fall in that group.

### Prioritizing Accessions

The user may click the “Prioritize Accessions” button to have all accessions that meet the default standards selected, as designated by the left-hand checkbox in the Level 1 Report table. The default settings prioritize protein Accessions that begin: - NP\_, O\_, P\_, Q\_, ABC\_, XP\_ and deselect those that are annotated as Other Sequences, unidentified, Unclassified sequence, Bacteria, unnamed, uncharacterized, partial, hypothetical, LOW QUALITY PROTEIN, unknown, green-fluorescent protein, as these proteins typically do not provide useful results in Level 3. The user may choose to manually

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

update the priority selections by adding and removing accessions using the checkboxes in the first column of the Level 1 Data - Primary Report table at the bottom of the Level 1 page. After selecting “Prioritize Accessions”, the number of accessions prioritized by the user will be displayed below the “Prioritize Accessions” button with the “x species selected” text. Prioritized accessions will be aligned in the Level 3 Analysis upon selecting “Request Residue Run”.

The screenshot shows a web form titled "Level 3 Query Amino Acid Residues". It contains several input fields and a dropdown menu. The fields are: "Select Template Sequence" (empty), "Additional Comparisons (optional)" (empty), "Enter Level 3 Run Name" (containing "Actinopteri"), and "Choose Taxonomic Group(s)" (containing "All Groups"). A dropdown menu is open, showing a list of taxonomic groups: "All Groups", "Actinopteri", "Amphibia", "Anthozoa", "Appendicularia", "Arachnida", and "Ascidiacea". At the bottom of the form, there are two buttons: "View Combined Report" and "Combine Level 3 Data".

**Note:** The “Choose Taxonomic Group(s):” drop-down will display the level of the taxonomic hierarchy being displayed in the “Filtered Taxonomic Group” column of the “Level 1 Data” table. For example, if the user changes the default option from “class” to “order,” then “order will be displayed in the dropdown.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

The screenshot shows the 'Level 3' configuration window. At the top is a 'Reference Explorer' button. Below it is the 'Level 3 Query Amino Acid Residues' section, which includes a 'Select Template Sequence' field containing 'NP\_000116.2' and an 'Additional Comparisons (optional)' field. The 'NCBI COBALT' section has an 'Enter Level 3 Run Name' field with 'Order not Class' entered. The 'Choose Taxonomic Group(s)' dropdown menu is open, displaying a list of taxonomic groups: 'All Groups', 'Acipenseriformes', 'Actiniaria', 'Amphipoda', 'Anabantiformes', 'Anguilliformes', and 'Anseriformes'. At the bottom of the window are 'View Combined Report' and 'Combine Level 3 Data' buttons.

By choosing a group from the drop-down menu, the “Level 1 Data” table below will be filtered by the selected Taxonomic Group (see column “Taxonomic Group” in “Level 1 Data” table). When a “Taxonomic Group” is selected from the drop-down, it can take up to a few seconds for the “Level 1 Data” table to filter completely, depending on the size of the table. The user can then examine each hit protein in the “Level 1 Data” table and select those that they would like to compare to the template sequence. To select sequences/species from the filtered “Level 1 Data” table, the user will select the check boxes in the first column of the table. Although it is not typically recommended, the user may also select the header check box in the first column to select all sequences/species in the filtered table.

**Note:** The user can also type the “Taxonomic Group” of interest in the text search box at the top of the drop-down for quick filtering.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Below is an example where the user selected the “Taxonomic Group” Actinopteri from the drop-down and then selected individual sequences/species to align with the template sequence. The number of selected species will be shown in the text above the “Request Residue Run” button.

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name	BLASTp Bitscore	Ortholog Candidate	Ortholog Count
6	MEN330156.1	16743	7324	Actinopteri	Actinopteri	<i>Amia calva</i>	Bowfin	ESR1 protein	647.51	Y	655
6	BAG82653.1	94	812342	Actinopteri	Actinopteri	<i>Atractosteus tropicus</i>	Tropical gar	estrogen_receptor_alpha	645.97	Y	658
6	MEN3320118.1	15567	7317	Actinopteri	Actinopteri	<i>Atractosteus spatula</i>	Alligator gar	ESR1 protein	644.04	Y	656
6	XP_006624508.1	41063	7318	Actinopteri	Actinopteri	<i>Leiocassis ocellatus</i>	Spotted gar	PREDICTED_estrogen_receptor	641.34	Y	656
6	BXM14939.1	89392	7306	Actinopteri	Actinopteri	<i>Acipenser ruberatus</i>	Sterlet	Estrogen receptor	629.79	Y	658
6	MEN3261383.1	18732	7313	Actinopteri	Actinopteri	<i>Esovelodon roathui</i>	Mississippi paddlefish	ESR1 protein	600.51	Y	658
6	BAG82650.1	250	111384	Actinopteri	Actinopteri	<i>Acipenser schrenckii</i>	Amur sturgeon	estrogen_receptor_alpha1	595.12	Y	658

(See **Search, View, and Download Data Tables** section of user guide for more information)

The user can choose to align sequences/species from multiple taxonomic groups with the template sequence, by going back to the “Choose Taxonomic Group” drop-down and selecting another group, which filters the Level 1 table based on the group selected, and then the user can select additional species from the newly filtered table. As before, the number of selected species can be tracked in the text above the “Request Residue Run” button that reads “X species selected”.

When the user has selected all sequences they want to align, then click the “Request Residue Run” button. Upon successful submission of a Level 3 query the user will see the following pop-up message. If submission is unsuccessful, a message will appear describing the reason for the unsuccessful submission.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

To update the “Choose Query to View” drop-down menu with the completed Level 3 alignments, the user can click on the “Refresh Level 2 and 3 runs” button.

Additionally, the user can check the status of the Level 3 run by clicking the “SeqAPASS Run Status” tab and the radio button for “Level 3 Status.” Typically, Level 3 alignments complete in a few seconds. When the Level 3 query completes and the Level 1 page has been updated, the user defined Level 3 Run Name will be available in the “Choose Query to View” drop-down menu. After selecting the desired Run Name from the drop-down, click “View Level 3 Data” button to view the aligned sequences and set up the individual amino acid residue alignments with the selected sequences/species.

Upon a successful Level 3 query submission a pop-up message will be displayed as follows in the upper right-hand side of the screen:

Once the Level 3 run has completed, the user can select the “Select Level 3 Run Name” drop down in the “View Single Report” box to view an individual user defined Level 3 run. If the user has completed multiple Level 3 alignments, between a template sequence and more than one taxonomic groups, the user can combine Level 3 reports by selecting the “Combine Level 3 Data” button. A pop-up will appear for the “Combine Level 3 Reports”. There are a series of three steps to combine Level 3 reports. First the user

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

will “Choose a Level 3 Template” from the dropdown that contains a list of all templates used to generate alignments in Level 3 by the user. The template sequence must be in-common to the Level 3 runs that will be combined.



After selecting the template, the user will click the “Next” button. At this point the user will select all Level 3 Jobs that are to be combined by selecting the check box in the “Level 3 Jobs” dropdown next to the user defined names. After all jobs that are to be combined are selected the user will click the “Next” button. Note that as the user moves through each step of the Combine Level 3 Reports feature, the step the user is currently on is indicated by highlighting the button in blue coloring (example “Level 3 Jobs” button is highlighted when working on selecting Jobs to combine).



The next step in the “Combine Level 3 Reports” feature is to put the jobs in order as to how they should be displayed in the output. Typically, sequences from an individual taxonomic group are aligned to a template sequence and named accordingly (e.g., Actinopteri, Amphibia, Aves, etc.). It may be useful to order the combined report similarly to how the taxonomic groups are displayed on the x-axis of the Level 1 or Level 2 data visualization. Therefore, the user can select the user defined name from the “Order Level 3 Jobs:” text box and drag and drop the name to the desired order from top to bottom. To move on to select individual amino acids for sequence comparisons the user will select the “View Level 3 Data” button.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov



The order selected will translate to the top to bottom order displayed in the data table, with the template sequence only displayed once in the first row and all selected jobs below.

**Level 3 Data - Primary**

The following links exit the site: [EXIT](#) [Download Current Level 3 Report Settings](#)

Search:

Data Version	Job Name	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Scientific Name
4	Amphibia	<a href="#">NP_000116.2</a>	1265506	<a href="#">9606</a>	Mammalia	<a href="#">Homo sapiens</a>
4	Amphibia	<a href="#">OCT77903.1</a>	130454	<a href="#">8355</a>	Amphibia	<a href="#">Xenopus laevis</a>
4	Amphibia	<a href="#">BAE30926.1</a>	83	<a href="#">186789</a>	Amphibia	<a href="#">Andrias japonicus</a>
4	Amphibia	<a href="#">AUW64608.1</a>	1591	<a href="#">141262</a>	Amphibia	<a href="#">Andrias davidianus</a>
4	Amphibia	<a href="#">BAE81788.1</a>	94392	<a href="#">8364</a>	Amphibia	<a href="#">Xenopus tropicalis</a>
4	Amphibia	<a href="#">BAJ05031.1</a>	18	<a href="#">2040589</a>	Amphibia	<a href="#">Sclerophrys capensis</a>
4	Aves	<a href="#">XP_019468458.1</a>	34219	<a href="#">9103</a>	Aves	<a href="#">Meleagris gallopavo</a>
4	Aves	<a href="#">XP_025878017.1</a>	31563	<a href="#">8790</a>	Aves	<a href="#">Dromaius novaehollandiae</a>
4	Aves	<a href="#">KFQ02396.1</a>	30590	<a href="#">8989</a>	Aves	<a href="#">Haliaeetus albicilla</a>
4	Aves	<a href="#">XP_010580195.1</a>	25311	<a href="#">52844</a>	Aves	<a href="#">Haliaeetus leucocephalus</a>

(1 of 2) 1 2 10 Download Table:



## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

View, and Download Data Tables section of user guide for more information.

For additional information on Amino Acid Residues, including definition of the acronym, the amino acid residue name, the classification for the amino acid side chain and the size of the amino acid residue based on molecular weight, the user can click the “Show Amino Acid Info...” button. A pop-up table, “Amino Acid info,” will be displayed providing this information.

The screenshot shows the 'Level 3 Template Protein Information' interface. A pop-up window titled 'Amino Acid info' is open, displaying a table of amino acid properties. The table has columns for ID, Name, Side Chain, and Size. The main interface shows a list of amino acid residues (1M, 2T, 3M, 4T, 5L, 6H, 7T, 8K, 9A) and an 'Update Report' button.

ID	Name	Side Chain	Size
A	Alanine	Aliphatic	89.094
C	Cysteine	Sulfur-Containing	121.154
D	Aspartic Acid	Acidic	133.104
E	Glutamic Acid	Acidic	147.131
F	Phenylalanine	Aromatic	165.192
G	Glycine	Aliphatic	75.067
H	Histidine	Basic	155.150
I	Isoleucine	Aliphatic	131.175
K	Lysine	Basic	146.189
L	Leucine	Aliphatic	131.175
M	Methionine	Sulfur-Containing	149.208
N	Asparagine	Amidic	132.119
P	Proline	Aliphatic	115.132
Q	Glutamine	Amidic	146.148
R	Arginine	Basic	174.203
S	Serine	Hydroxylic	105.093
T	Threonine	Hydroxylic	119.119
U	Seleno-cysteine	Sulfur-Containing	168.054
V	Valine	Aliphatic	117.148
W	Tryptophan	Aromatic	204.228
X	Unknown	Unknown	-
Y	Tyrosine	Aromatic	181.191

To obtain individual amino acid residue alignment data in the Level 3 data table, the user must use the shuttle in the “Level 3 Template Protein Information” box to select positions and amino acid residues from the chosen template sequence to align with the sequences/species that were selected by taxonomic group. Single letter abbreviations are used for the amino acid sequences.

G: Glycine    A: Alanine    S: Serine    T: Threonine    C: Cysteine    V: Valine  
L: Leucine    I: Isoleucine    M: Methionine    P: Proline    F: Phenylalanine    U: Seleno-cysteine  
Y: Tyrosine    W: Tryptophan    D: Aspartic Acid    E: Glutamic Acid  
N: Asparagine    Q: Glutamine    H: Histidine    K: Lysine    R: Arginine

The screenshot shows the 'Select Amino Acid Residues' interface. It features two boxes: a left-hand box with a list of residues (1M, 2T, 4T, 5L, 6H, 7T, 8K, 9A) and a right-hand box with a list of residues (3M, 219Y, 267H, 268K, 272D, 594T). A shuttle with arrows is positioned between the boxes. An 'Update Report' button is located below the left-hand box.

The user can select one residue at a time by clicking and highlighting the residue of interest and then clicking the top right arrow shuttle button to move the residue to the right-hand box for inclusion in the alignment. Each time a residue is added to the right-hand box, the left-hand box resets itself to the 1<sup>st</sup> residue. Or the user can select multiple residues at the same time by holding the Ctrl button, clicking on residues, and then clicking the top right arrow shuttle button to move the residues to the right-hand box. The user can choose to remove selected residues by using the left arrow button to clear one at a time or the double left arrow button to remove all selected residues at once. When residues of interest (likely

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

defined from the literature as described above) have been selected, click the “Update Report” button, which then updates the “Level 3 Data” table with the individual residue alignment data.

Alternatively, the user can enter the amino acid positions in the “Enter Amino Acid Residue Positions” text box (e.g., 351,353,362,364,394,524) and click the “Copy to Residue List” button.

Upon clicking “Copy to Residue List” the “Select Amino Acid Residues” shuttle box is populated with the position and residues typed. The user can then click the update Report button to produce Level 3 results in the table below.

The screenshot displays the SeqAPASS interface. At the top, there is a text input field labeled "Enter Amino Acid Residue Positions" containing the text "351,353,362,364,394,524". Below this field is a blue button labeled "Copy to Residue List". Below the button is a window titled "Select Amino Acid Residues". This window contains two lists of amino acid residues. The left list shows positions 1M through 10C. The right list shows residues 351D, 353E, 362K, 364V, 394R, and 524H. Between the lists are three buttons: a right-pointing arrow, a left-pointing arrow, and a double-headed arrow. To the right of the lists is another "Enter Amino Acid Residue Positions" field containing the same text, and a "Copy to Residue List" button. At the bottom left of the window is an "Update Report" button.

The individual amino acid residue alignment data will then be updated on the right most columns of the Level 3 Data table. The user *can submit a maximum of 50 individual amino acid residues* from the template sequence to compare to the other selected sequences. The individual amino acid residues will be listed in numerical order starting with the 1<sup>st</sup> position in the template sequence to the last position in the template sequence.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Level 3 Data – Primary Report

The default report is the “Primary Report” and can be recognized as such because the radio button for “Primary Report” above the “Level 3 Data” table is selected.

The “Primary Report” columns for the alignment will be titled “Similar Susceptibility as Template” (“Y” or “N” for yes or no, respectively), followed by Position 1, Amino Acid 1, Total Match 1, Position 2 Amino Acid 2, Total Match 2, Position 3, Amino Acid 3, Total Match 3.... The template sequence will always be in the top row of the “Level 3 Data” table followed by the previously selected sequences. Further, the residues selected in the shuttle will also be displayed in the top row corresponding to the template sequence. Each Position and Amino Acid in the following rows are those corresponding to the Protein Accession identified in that row and aligning with the template sequence. The Total Match X describes whether the amino acid residue matches the template based on side-chain classification and molecular weight, “Y,” for yes, or “N,” for not a match to the template. The user can evaluate this data to understand how well conserved an amino acid residue is across species or in a species of interest to add an additional line of evidence to support (or question) susceptibility predictions. The user can also download the current report settings by selecting the “Download Current Level 3 Report Settings.” This csv allows the user to track which settings were used or changed by the user when downloading a data table.

Protein Name	Analysis Completed	Similar Susceptibility as Template	Position 1	Amino Acid 1	Total Match 1	Position 2	Amino Acid 2	Total Match 2	Position 3	Amino Acid 3	Total Match 3
<a href="#">estrogen receptor isoform 1</a>	2019 08 29 14:55:59	Y	351	D	Y	353	E	Y	362	K	Y
<a href="#">estrogen receptor alpha</a>	2019 08 29 14:55:59	Y	320	D	Y	322	E	Y	331	K	Y
<a href="#">PREDICTED_estrogen receptor isoform X2</a>	2019 08 29 14:55:59	Y	316	D	Y	318	E	Y	327	K	Y
<a href="#">estrogen receptor</a>	2019 08 29 14:55:59	Y	355	D	Y	357	E	Y	366	K	Y
<a href="#">estrogen receptor isoform X3</a>	2019 08 29 14:55:59	Y	319	D	Y	321	E	Y	330	K	Y
<a href="#">Estrogen receptor 1</a>	2019 08 29 14:55:59	Y	319	D	Y	321	E	Y	330	K	Y

When downloading the current “Level 3 Report Settings”, the following information will be present in the csv. If the user decides to change the default settings, the csv can be utilized for quick information if the SeqAPASS page is no longer open.

	A	B
1	Level 3 Report Settings	
2		
3		
4	Analysis TimeStamp	2019 05 16 11:04:08
5	SeqAPASS version	3.2
6	Level 3 Run Name	Actinopteri
7	Template Species	Homo sapiens
8	Template Protein	[NP_000116.2] estrogen receptor isoform 1
9	Query Residues	1M, 2T, 3M, 4T, 5L, 6H, 7T, 8K, 9A, 10S
10	Query Accession	NP_000116.2

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Level 3 Data – Full Report

The user may choose to view the Full Report for Level 3 data by selecting the radio button above the “Level 3 Data” table for “Full Report.” The table below will automatically update to display all of the alignment details.

The “Full Report” columns for the alignment will be titled “Similar Susceptibility as Template” (“Y” or “N” for yes or no respectively), followed by Position 1, Amino Acid 1, Direct Match 1, Side Chain 1, MW1, MW Match 1, Total Match 1, Position 2, Amino Acid 2, Direct Match 2, Side Chain 2, MW2, MW Match 2, Total Match 2..... The template sequence will always be in the top row of the “Level 3 Data” table followed by the previously selected sequences. Further, the residues selected in the shuttle will also be displayed in the top row corresponding to the template sequence. Each Position and Amino Acid in the following rows are those corresponding to the Protein Accession identified in that row align with the template sequence. The Total Match X describes whether the amino acid residue matches the template based on side-chain classification and molecular weight, “Y,” for yes, or “N,” for not a match to the template. The user can evaluate this data to understand how well conserved an amino acid residue is across species or in a species of interest to add an additional line of evidence to support (or question) susceptibility predictions.

The following links exit the site: [EXIT](#) [Download Current Level 3 Report Settings](#)

Search:

Analysis Completed	Similar Susceptibility as Template	Position 1	Amino Acid 1	Direct Match 1	Side Chain 1	Side Chain Match 1	MW 1	MW Match 1	Total Match 1	Position 2	Amino Acid 2
2019 08 29 14:55:59	Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E
2019 08 29 14:55:59	Y	320	D	Y	Acidic	Y	133.104	Y	Y	322	E
2019 08 29 14:55:59	Y	316	D	Y	Acidic	Y	133.104	Y	Y	318	E
2019 08 29 14:55:59	Y	355	D	Y	Acidic	Y	133.104	Y	Y	357	E
2019 08 29 14:55:59	Y	319	D	Y	Acidic	Y	133.104	Y	Y	321	E
2019 08 29 14:55:59	Y	319	D	Y	Acidic	Y	133.104	Y	Y	321	E

(1 of 1) [Download Table:](#)

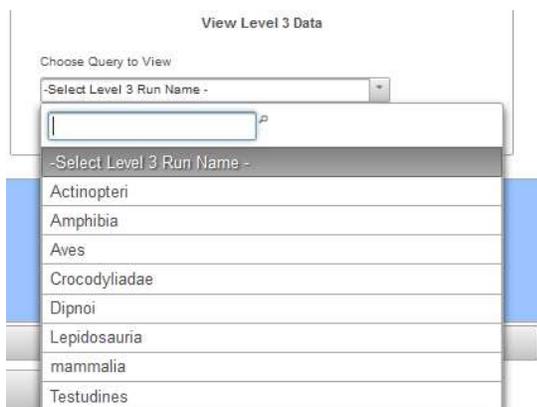
The “Direct Match X” column describes whether the hit amino acid is an exact match to the template amino acid, providing a “Y” or “N” for yes or no, respectively. The “Side Chain X” column indicates the side chain classification for the amino acid residue (click on “Show Amino Acid Info...for more information on classifications). The “Side Chain Match X” column indicates whether the hit side chain has the same classification as the template amino acid, providing a “Y” or “N” for yes or no, respectively. The “MW X” column indicates the molecular weight (g/mol) of the amino acid residue and the “MW Match X” column indicates whether the hit molecular weight has a difference in molecular weight greater than or equal to 30 g/mol compared to the template amino acid, providing a “Y” or “N” for yes or no, respectively. For the “Total Match X” to be “Y,” both “Side Chain Match X” and “MW Match X” should be either “Y” and “Y” or one “Y” and one “N,” respectively. Only if both “Side Chain Match X” and “MW Match X” are “N” and “N,” then the “Total Match X” is “N” for no. Ultimately, the Total Match 1, 2, 3, 4... are used to inform the “Similar Susceptibility as Template” column. If there is one or more “N” for Total Match comparing any amino acid residue to the template across a row for a given species, then the “Similar Susceptibility as Template” is “N” for no, indicating that the hit species is predicted NOT to have the same susceptibility prediction as the template sequence. However, if all “Total Match X” are “Y” for yes, then the “Similar Susceptibility as Template” is “Y” indicating that the hit species is predicted to have the same susceptibility prediction as the template sequence.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

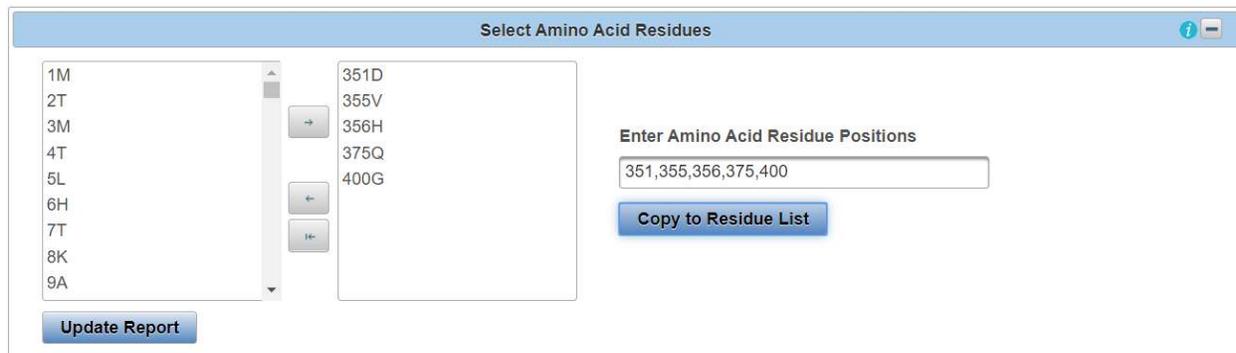
Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Multiple Level 3 Runs Requiring the Same Amino Acid Residue Comparisons

Typically, Level 3 individual amino acid residue alignments are submitted repetitively, comparing species from one taxonomic group at a time to the template amino acid residue(s).



Therefore, to increase efficiency in submitting the same alignments in Level 3 over and over again, the user can take advantage of the “Copy to Residue List” button. For the first alignment of amino acid residues, the user would select the amino acid residues to align and click the “Update Report” button.



By clicking “Update Report” the residues that were selected will be copied into the “Enter Amino Acid Residue Positions” text box. When the user selects a new “Level 3 Run Name” (from the same Level 1 query accession) to view by using the “View Level 3 Data” dropdown and clicking the “View Level 3 Data” button on the “Level 1 Query Protein Information” page, the “Enter Amino Acid Residue Positions” text box will be populated with the amino acid residues selected from the previous run.



The user can keep, add, or delete, residue positions in this box and click “Copy to Residue List” button. The amino acid residues will then be moved to the “Select Amino Acid Residues Shuttle” and the user can then click “Update Report” to view the data in the table below.

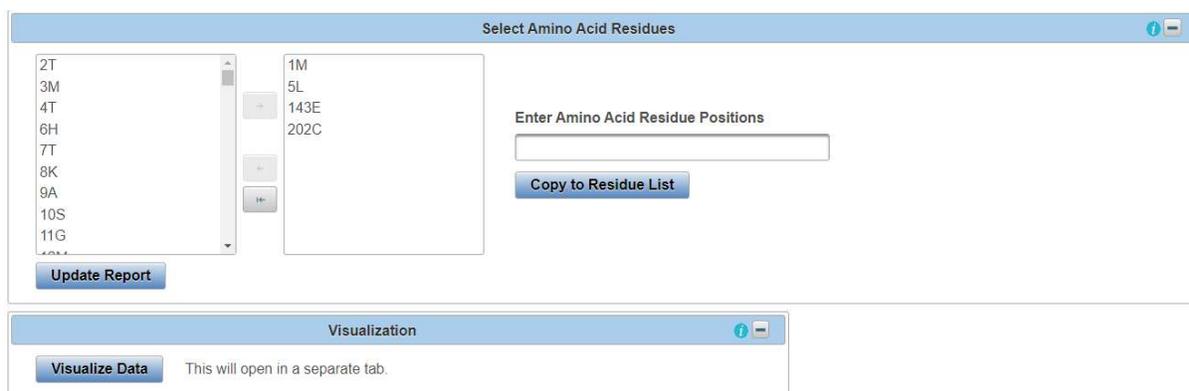
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

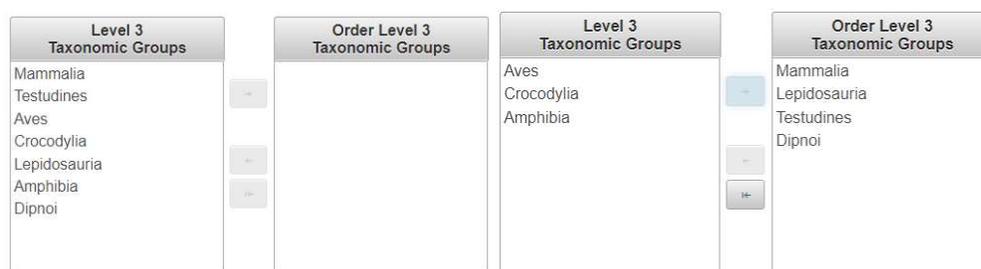
### Heat Map

The Heat Map is a feature that allows the user to have a visual representation of the chosen amino acid(s) for a single Level 3 run. The Heat Map utilizes color to denote which amino acids are a total match, partial match and not a match to the template sequence. The Heat Map is accessed within the “Level 3” page under the “Visualization” drop down and will open up in a separate tab. The Heat Map has many similar features to the Level 1 and 2 boxplot with some added customizable features. There are many settings that can be changed within the Heat Map and if necessary, there are informational buttons that can be opened to get added information regarding the different options.

To get to the Heat Map, open a completed Level 3 run and click the “Visualization” drop down then select the “Visualize Data” button. This will bring you to the Heat Map where there is information regarding the features of the map. Then select the “Heat Map” icon to access the Heat Map itself.



The default order of the taxonomic groups is based on how the species are selected during the Level 3 set up process. There is the option to include all taxonomic groups or a user chosen few. To move the taxonomic groups over to place them in order you must either click or \*CTRL\* click and select the arrow pointed to the right. Once the taxonomic groups are moved over, the user can order the groups by dragging them up or down.



### Report Options

There are multiple options within the Heat Map that can be changed based on what information the user desires to have present. The Heat Map itself can be changed between the “Simple” report which shows the amino acid and its respective position or the “Full” report which gives added information about each amino acid. The user can also change between the common name and scientific name displayed on the Heat Map.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Optional Selections

The “Optional Selections” for the Heat Map will highlight the name for each respective species based on what is selected; Ortholog Candidates, Threatened Species, Endangered Species, Common Model Organism. Only one optional selection can be highlighted at a time.

Common Name	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4
Human	32K	46S	55P	64A
Diamondback terrapin	32K	46S	55P	64T
Western painted turtle	32K	46S	55P	64T
Chinese soft-shelled turtle	32K	46S	55P	64T
Terrapins	32K	46S	55P	64T
Goodes thornscrub tortoise	32K	46S	55P	64T
Pacific ridley	32K	46S	55P	64T
Painted turtle	32K	46S	55P	64T
Green sea turtle	32K	46S	55P	64T
Three-toed box turtle	61K	75S	84P	93T

## Heat Map Settings

Changing the “Heat Map Settings” will give the user the option to display specific information in the Heat Map. The user can select or deselect a variety of the settings to have a customized Heat Map.

**Report Options** i -

**Report Type**

Simple

Full

**Species Name Type**

Common Name

Scientific Name

**Optional Selections** i -

Ortholog Candidates  Threatened Species  Endangered Species  Common Model Organisms

**Heat Map Settings** i -

Susceptibility Prediction Heat Map

Susceptibility Prediction Text

Alignment Prediction Heat Map

Amino Acid

Amino Acid Position

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Common Name	Similar Susceptibility	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4
Human	Y	133E	135E	137S	584I
Western painted turtle	Y	127E	129D	131S	577V
Lappet-faced vulture	N	125E	127E	129G	575I
Nile crocodile	Y	127E	129D	131S	575I
Split-tongued squamates	Y	127E	129E	131N	576V
Japanese giant salamander	N	126E	128E	130G	573L
West African lungfish	Y	131E	133E	135S	580G

■ Total Match  
■ Partial Match  
■ Not a Match

Above is an example of a simple report which shows the amino acid and its respective position. Each amino acid is compared to the template species and can receive a dark blue color (Total match), a light blue color (Partial match), or a yellow color (Not a match). To access more information regarding each amino acid, the user can scroll over the amino acid box to bring up a box with added data.

How amino acids are compared to the template: Comparing Side Chain Classification (acidic, basic, aromatic, etc.) and Molecular weight as surrogate for size (> 30g/mol different in size). Both the same (total match), One the same (partial match), Both differ (not a match).

Below is an example of a full report which also shows the amino acid and its respective position but also shows the amino acid's side chain, molecular weight and if it is a Total match (dark blue) or Not a match (yellow) to the template species.

Common Name	Similar Susceptibility	Amino Acid 1	Side Chain 1	MW 1	Total Match 1	Amino Acid 2	Side Chain 2	MW 2	Total Match 2	Amino Acid 3	Side Chain 3	MW 3	Total Match 3
Human	Y	274G	Aliphatic	75.067	Y	275E	Acidic	147.131	Y	276G	Aliphatic	75.067	Y
Western painted turtle	N	268Q	Amidic	146.146	N	269D	Acidic	133.104	Y	270A	Aliphatic	89.094	Y
Nile crocodile	N	268Q	Amidic	146.146	N	269D	Acidic	133.104	Y	270A	Aliphatic	89.094	Y
Split-tongued squamates	N	268Q	Amidic	146.146	N	269D	Acidic	133.104	Y	270S	Hydroxylic	105.093	N
Japanese giant salamander	N	267P	Aliphatic	115.132	Y	268D	Acidic	133.104	Y	269Q	Amidic	146.146	N

■ Match  
■ Not a Match

The example below shows only the “Alignment Prediction” (Amino acid match against template) for each amino acid in chronological order.

**Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide**  
 Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Common Name	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4
Human	Dark Blue	Dark Blue	Dark Blue	Dark Blue
Western painted turtle	Dark Blue	Light Blue	Light Blue	Light Blue
Lappet-faced vulture	Dark Blue	Dark Blue	Yellow	Dark Blue
Nile crocodile	Dark Blue	Light Blue	Dark Blue	Dark Blue
Split-tongued squamates	Dark Blue	Dark Blue	Light Blue	Light Blue
Japanese giant salamander	Dark Blue	Dark Blue	Yellow	Light Blue
West African lungfish	Dark Blue	Dark Blue	Dark Blue	Light Blue

■ Total Match  
■ Partial Match  
■ Not a Match

There is added information for each species (NCBI Accession, Protein Name, Scientific Name, and Taxonomic Group) along with each amino acid (Amino Acid Name, Abreviation, Side Chain, and Molecular Weight). This can be found by scrolling over the species name or the amino acid.

Common Name	Similar Susceptibility	Amino Acid 1	Amino Acid 2								
Human	Y	274G	275E								
Western painted turtle	<table border="1"> <tr> <td>NCBI Accession</td> <td>NP_000116.2</td> </tr> <tr> <td>Protein Name</td> <td>estrogen receptor isoform 1</td> </tr> <tr> <td>Scientific Name</td> <td>Homo sapiens</td> </tr> <tr> <td>Taxonomic Group</td> <td>Mammalia</td> </tr> </table> Ortholog Candidate	NCBI Accession	NP_000116.2	Protein Name	estrogen receptor isoform 1	Scientific Name	Homo sapiens	Taxonomic Group	Mammalia		
NCBI Accession		NP_000116.2									
Protein Name		estrogen receptor isoform 1									
Scientific Name		Homo sapiens									
Taxonomic Group	Mammalia										
Nile crocodile											
Split-tongued squamates											
Japanese giant salamander											

■ Total Match  
■ Partial Match  
■ Not a Match

ty	Amino Acid 1	Side Chain 1	MW 1	Total Match 1
	274G	Aliphatic	75.067	Y
	268Q	Name	Glycine	N
	268Q	Abv	G	N
	268Q	Side Chain	Aliphatic	N
	267P	MW	75.067	Y

To push the designed Heat Map to the Decision Summary Report as a visualization, press the “Push Level 3 Heatmap to DS Report” button. It will then be active within the DS Report Level 3 section. To

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

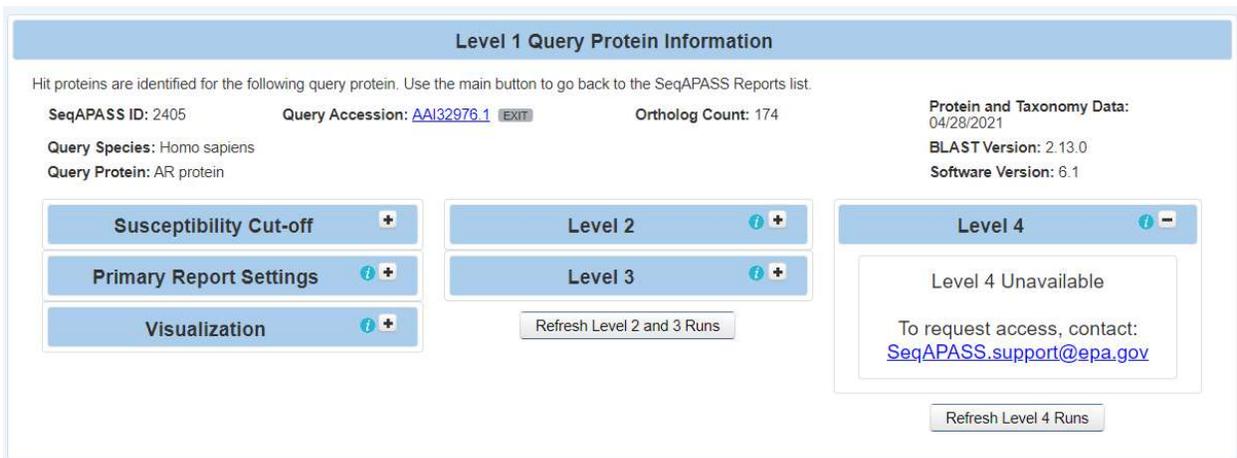
Updated 08/14/2023; Contact Carlie LaLone with Questions: [LaLone.Carlie@epa.gov](mailto:LaLone.Carlie@epa.gov)

download the Heat Map, press the “Download Heatmap...” button. The Heat Map can be downloaded as an SVG, JPG or PNG.

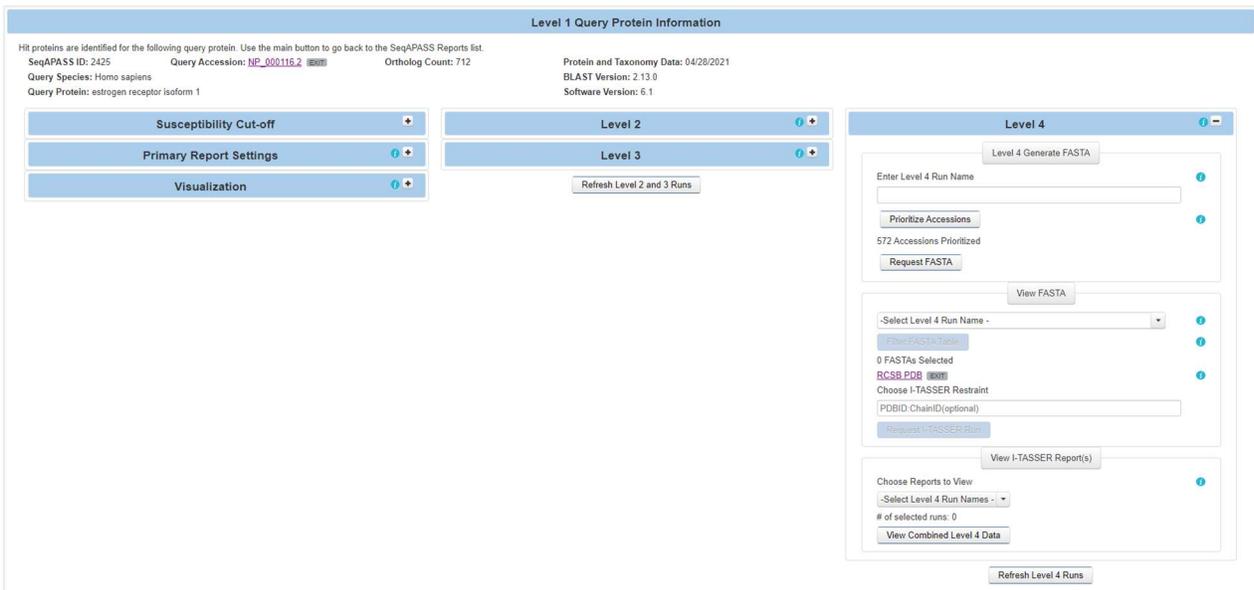


## Level 4: Protein Structural Analysis

Experienced users may access Level 4 to utilize the integrated protein structural analysis. Level 4 access can be requested through the SeqAPASS Homepage “Submit a Problem/Question” link. If a user does not have Level 4 access, expanding the Level 4 tab will display the message “Level 4 Unavailable. To request access, contact [SeqAPASS.support@epa.gov](mailto:SeqAPASS.support@epa.gov)”.



The Iterative Threading ASSEMBly Refinement (I-TASSER), AlphaFold, and TM-align tools. The I-TASSER and AlphaFold tools are capable of 3D structural generation, while TM-align analyses multiple structures for similarity as another line of evidence toward conservation.

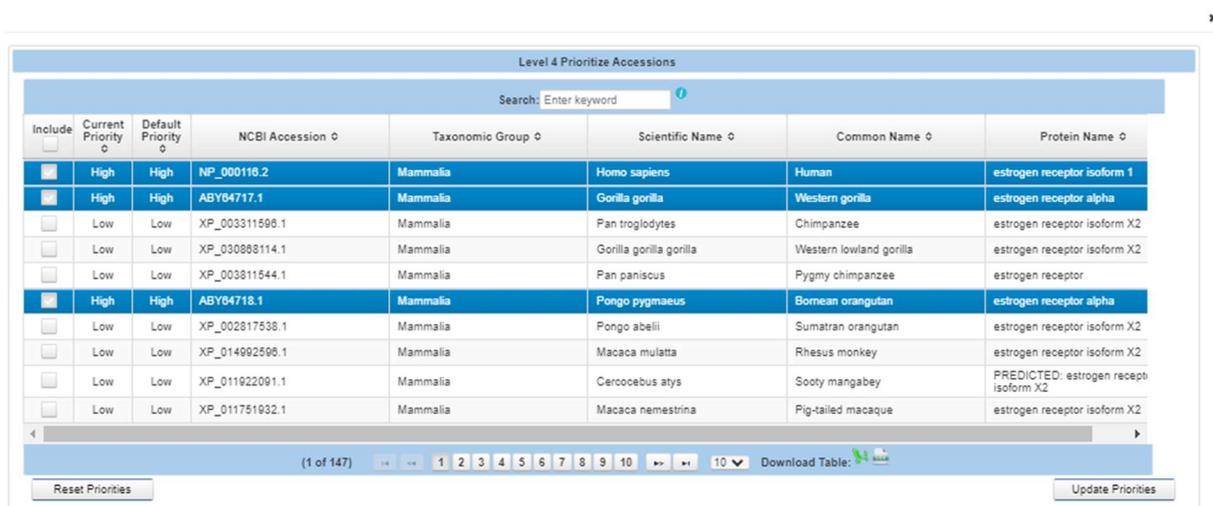


## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Level 4 FASTA Generation

To begin the Level 4 analysis, the NCBI accessions from the Level 1 alignments must be converted into FASTA formatted sequences for submission to I-TASSER. Users may prioritize accessions for FASTA generation in the “Level 4 Generate FASTA” subsection by providing a user defined name for their job in the “Enter Level 4 Run Name” text box. After naming the run, the user may prioritize accessions for FASTA generation through selecting the “Prioritize Accessions” button where a “Level 4 Prioritize Accessions” table will pop up. This table will, by default, have likely high-quality accessions preselected for FASTA generation, however, the user may add or remove any accessions available in the table using the selection checkboxes. After selecting and/or removing the desired accessions, the user may click the “Update Priorities” button to confirm their prioritizations or select the “Reset Priorities” to undo any changes they have made from the default prioritizations (The “x” button may also be utilized to close the “Level 4 Prioritize Accessions” table without committing any changes).



Include	Current Priority	Default Priority	NCBI Accession	Taxonomic Group	Scientific Name	Common Name	Protein Name
<input checked="" type="checkbox"/>	High	High	NP_000116.2	Mammalia	Homo sapiens	Human	estrogen receptor isoform 1
<input checked="" type="checkbox"/>	High	High	ABY64717.1	Mammalia	Gorilla gorilla	Western gorilla	estrogen receptor alpha
<input type="checkbox"/>	Low	Low	XP_003311598.1	Mammalia	Pan troglodytes	Chimpanzee	estrogen receptor isoform X2
<input type="checkbox"/>	Low	Low	XP_030888114.1	Mammalia	Gorilla gorilla gorilla	Western lowland gorilla	estrogen receptor isoform X2
<input type="checkbox"/>	Low	Low	XP_003811544.1	Mammalia	Pan paniscus	Pygmy chimpanzee	estrogen receptor
<input checked="" type="checkbox"/>	High	High	ABY64718.1	Mammalia	Pongo pygmaeus	Bornean orangutan	estrogen receptor alpha
<input type="checkbox"/>	Low	Low	XP_002817538.1	Mammalia	Pongo abelii	Sumatran orangutan	estrogen receptor isoform X2
<input type="checkbox"/>	Low	Low	XP_014992598.1	Mammalia	Macaca mulatta	Rhesus monkey	estrogen receptor isoform X2
<input type="checkbox"/>	Low	Low	XP_011922091.1	Mammalia	Cercocebus atys	Sooty mangabey	PREDICTED: estrogen recept isoform X2
<input type="checkbox"/>	Low	Low	XP_011751932.1	Mammalia	Macaca nemestrina	Pig-tailed macaque	estrogen receptor isoform X2

The number of accessions selected in the “Level 4 Prioritize Accessions” table for FASTA conversion is reported below the “Prioritize Accessions” button within the “Level 4 Generate FASTA” subsection. This number will reflect any changes to the Level 4 “Prioritize Accessions” table that are submitted through the “Update Priorities” button.



Level 4

Level 4 Generate FASTA

Enter Level 4 Run Name

Testing

Prioritize Accessions

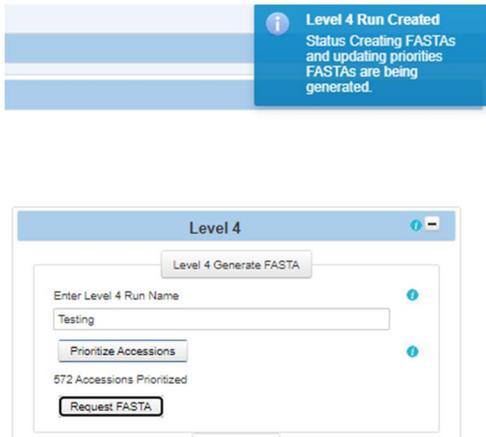
572 Accessions Prioritized

Request FASTA

When all the desired accessions are prioritized for FASTA generation, the “Request FASTA” button may be utilized to confirm the submission. A pop-up will appear in the upper right corner to notify a successful submission.

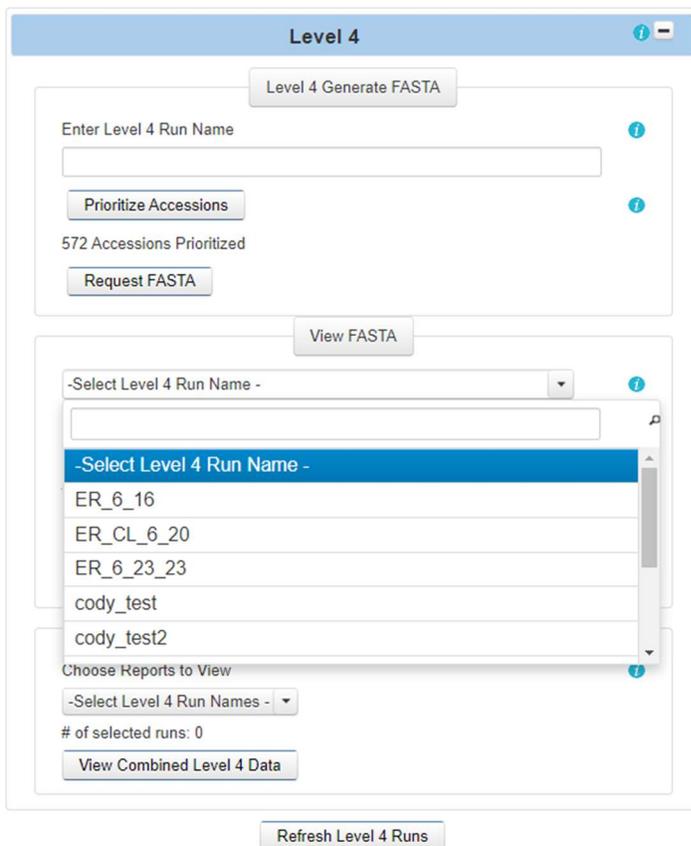
# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov



## View FASTA

After submitting accessions for FASTA generation, the user may select the “Refresh Level 4 Runs” button, located beneath the “Level 4” section, to refresh the available FASTA runs for selection. The user may then select a generated FASTA report to view by clicking the “Select Level 4 Run Name” dropdown menu and selecting the desired user defined name from the completed FASTA jobs.



## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

The user may then select the “Filter FASTA Table” button to bring up the “Level 4 Filter FASTAs” table. This table allows the user to add or remove generated FASTA sequences to push forward to I-TASSER structure generation (Since I-TASSER structure generation is a timely process, a maximum of 10 FASTAs may be submitted at a time. To submit additional FASTAs, a user will begin the process over in naming a job, generating FASTAs and prioritizing another batch of 10 jobs to submit to I-TASSER).

Include	Status	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name	BLASTp Bitscore	Ortholog Candidate	Cut-off	Percent Similarity	Susceptibility Prediction
<input checked="" type="checkbox"/>	FASTA created	AAW02952.1	47	283384	Insecta	Insecta	Melipona scutellaris	Stingless bees	ultraspiracle	168.777	N	34.43	13.43	N
<input checked="" type="checkbox"/>	FASTA created	NWQ96485.1	14023	240201	Aves	Aves	Burhinus bistriatus	Shorebirds and others	ESR1 protein	973.385	Y	34.43	78.38	Y
<input checked="" type="checkbox"/>	FASTA created	NWH24107.1	13830	9117	Aves	Aves	Grus americana	Whooping crane	ESR1 protein	970.304	Y	34.43	78.13	Y
<input type="checkbox"/>	FASTA created	NWQ75248.1	12327	115818	Aves	Aves	Columbina picui	Picui ground-dove	ESR1 protein	972.23	N	34.43	78.29	Y
<input type="checkbox"/>	FASTA created	ALB78124.1	130	9508	Mammalia	Mammalia	Ateles fusciceps	Brown-headed spider monkey	estrogen receptor beta	473.781	N	34.43	38.15	Y
<input type="checkbox"/>	FASTA created	NWW18570.1	14502	254539	Aves	Aves	Falco tinnunculus	Eastern shrikelet	ESR2 protein	473.011	N	34.43	38.09	Y

After adding and removing the desired FASTA accessions for I-TASSER, the user may select the “Update Changes” button to confirm their changes, or the “x” button to cancel their changes. After confirming any changes, the “x FASTAs Selected” text will display the current number of selected FASTAs for I-TASSERs submission.

View FASTA

Testing

Filter FASTA Table

3 FASTAs Selected

PDB Link EXIT

Choose I-TASSER Restraint

PDBID:ChainID(optional)

Request I-TASSER Run

The user may also choose a structural restraint for their I-TASSER structure generation by typing the PDBID:ChainID into the textbox beneath the “Choose I-TASSER Restraint” text. The I-TASSER Restraint serves a reference for I-TASSER’s structural generation.

View FASTA

-Select Level 4 Run Name -

Filter FASTA Table

0 FASTAs Selected

PDB Link EXIT

Choose I-TASSER Restraint

2AMA:A

Request I-TASSER Run

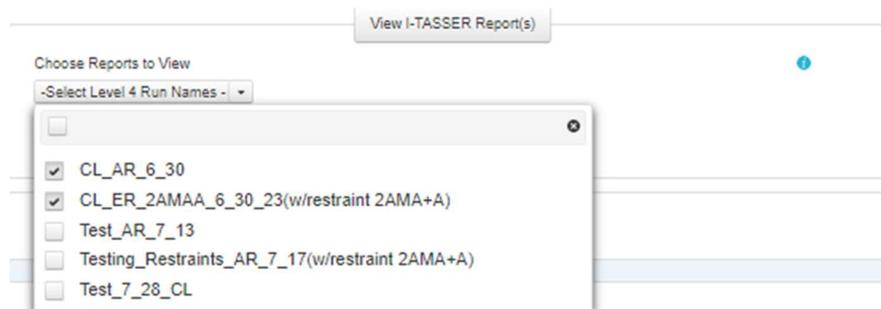
When all desired FASTAs are selected, the “Request I-TASSER Run” button may be clicked to submit the selected FASTAs to I-TASSER.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

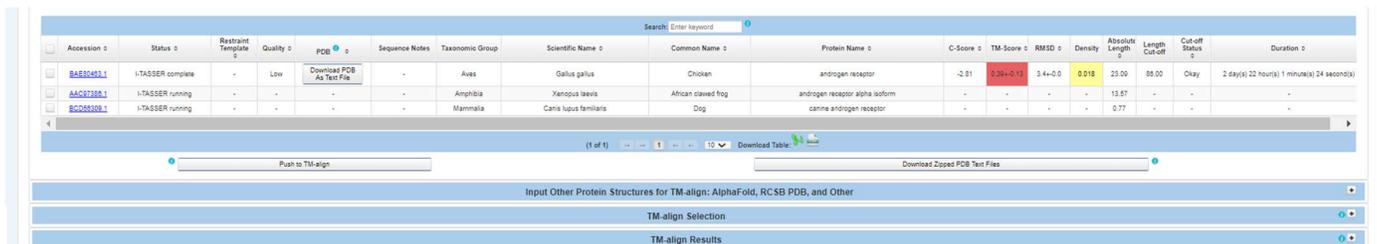
### View I-TASSER Report

After at least one job has completed from those submitted to I-TASSER, the user may select the “Refresh Level 4 Runs” button to reload the page with any available user defined report names. The user may then click the “Select Level 4 Run Names” dropdown menu and select their desired jobs for viewing. Multiple runs may be selected and combined. Since there is a maximum of 10 FASTAs per I-TASSER job, it is recommended that jobs of more than 10 FASTAs be split into multiple jobs, and combined later for viewing.



After selecting the desired job, the user may select the “View Level 4 Data” button to access the Level 4 Run Page. This section is for viewing singular I-TASSER jobs (which can contain multiple protein structures).

The Level 4 Data Page contains all available metrics from I-TASSER on the selected run including the C-Score, TM-Score, RMSD value, Density, Absolute Length, Length Cut-Off, Cut-Off Status, and Duration. The Table also reports taxonomic information associated with each accession, including the taxonomic group, scientific name, and common name.

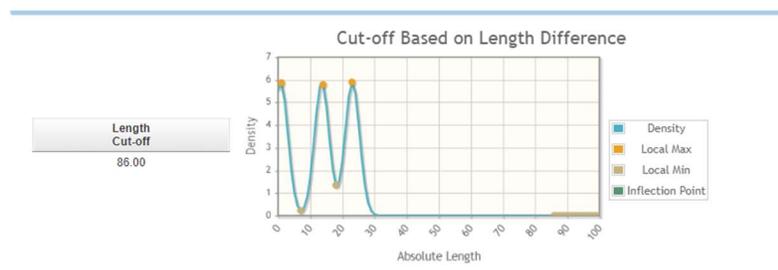


Accession	Status	Restrict Template	Quality	PDB	Sequence Notes	Taxonomic Group	Scientific Name	Common Name	Protein Name	C-Score	TM-Score	RMSD	Density	Absolute Length	Length Cut-Off	Cut-Off Status	Duration
BAG0401.1	I-TASSER complete	-	Low	Download PDB As Text File	-	Aves	Gallus gallus	Chicken	androgen receptor	-2.81	2.26-4.15	3.44-0.0	0.018	23.09	88.00	Okay	2 day(s) 22 hour(s) 1 minute(s) 24 second(s)
AAC0738.1	I-TASSER running	-	-	-	-	Amphibia	Xenopus laevis	African clawed frog	androgen receptor alpha isoform	-	-	-	-	13.57	-	-	-
BC058308.1	I-TASSER running	-	-	-	-	Mammalia	Canis lupus familiaris	Dog	canine androgen receptor	-	-	-	-	0.77	-	-	-

TM-align Selection

TM-align Results

The Level 4 Report page displays a length density plot providing the cut-off based on the absolute length difference between each hit structure and the query structure (defined by the Level 1 query sequence). The Length Cut-Off is determined by the second local minimum following the global maximum among the absolute lengths.



# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Level 4 Page

This page contains the completed I-TASSER job information, alongside a density plot displaying the length cut-offs and absolute lengths of the proteins. The “I-TASSER Result and PDB Selection” table displays all accessions with completed I-TASSER jobs. Each row of the data table contains the taxonomic group, scientific name, common name, and protein name, as well as I-TASSER data including the C-Score, TM-Score, RMSD value, Density, Absolute length, Length Cut-Off, Cut-Off Status, and I-TASSER Status.

Accession Notes	Taxonomic Group	Scientific Name	Common Name	Protein Name	C-Score	TM-Score	RMSD	Density	Absolute Length	Length Cut-off	Cut-off Status	Duration
-	Crocodylia	Alligator mississippiensis	American alligator	androgen receptor	-0.74	0.62+-0.14	4.6+-0.0	0.119	30.63	18.00	Check	2 day(s) 8 hour(s) 54 minute(s) 44 second(s)
-	Aves	Uria aalge	Common murre	ANDR protein	-0.86	0.61+-0.14	4.6+-0.0	0.127	21.66	18.00	Check	3 day(s) 20 hour(s) 40 minute(s) 30 second(s)
-	Testudinata	Platysternon megacephalum	Big-headed turtle	serine/arginine-rich splicing factor 5-like	-1.13	0.57+-0.14	4.6+-0.0	0.103	13.57	18.00	Okay	1 Week(s) 4 day(s) 21 hour(s) 36 minute(s) 33 second(s)
-	Actinopteri	Atractosteus spatula	Alligator gar	ANDR protein	-1.29	0.55+-0.15	4.5+-0.0	0.083	6.02	18.00	Okay	2 day(s) 23 hour(s) 23 minute(s) 48 second(s)
-	Amphibia	Xenopus laevis	African clawed frog	androgen receptor alpha isoform	-1.5	0.53+-0.15	4.4+-0.0	0.073	13.57	18.00	Okay	1 Week(s) 4 day(s) 21 hour(s) 36 minute(s) 33 second(s)
-	Aves	Uria aalge	Common murre	ANDR protein	-1.66	0.51+-0.15	4.4+-0.0	0.057	21.66	18.00	Check	1 Week(s) 4 day(s) 21 hour(s) 36 minute(s) 32 second(s)
-	Mammalia	Homo sapiens	Human	AR protein	-1.66	0.51+-0.15	4.4+-0.0	0.057	0.00	18.00	Okay	2794 Week(s) 5 day(s) 16 hour(s) 52 minute(s) 0 second(s)
-	Aves	Brachypteryx leptosomus	Short-legged ground-roller	ANDR protein	-2.63	0.41+-0.14	4.0+-0.0	0.017	46.06	18.00	Check	1 Week(s) 4 day(s) 21 hour(s) 36 minute(s) 33 second(s)
-	Testudinata	Pseudemys nelsoni	Terrapins	androgen receptor	-2.7	0.4+-0.14	3.4+-0.0	0.021	13.57	18.00	Okay	1 Week(s) 4 day(s) 21 hour(s) 36 minute(s) 33 second(s)
-	Aves	Urocolius indicus	Mousebirds	ANDR protein	-2.77	0.4+-0.13	3.8+-0.0	0.015	43.22	18.00	Check	1 day(s) 15 hour(s) 16 minute(s) 18 second(s)

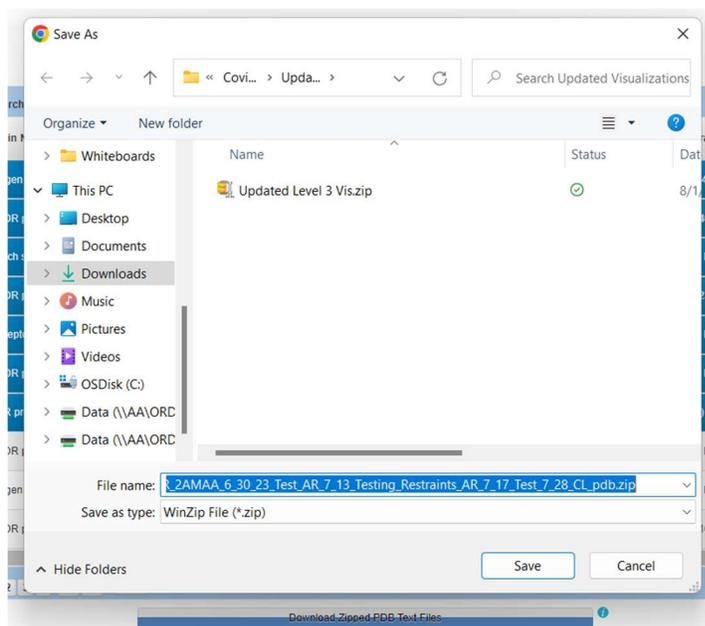
The completed I-TASSER job will be downloadable as a PDB formatted file in the table. If the Cut-off Status is “Check” it is recommended that the user use an external structure visualization tool to determine if the structure needs modification for further analysis. As an example, this can occur with nuclear receptors where some structures may be generated from sequences representing the full protein and others for the ligand binding domain only. The user is notified to “check” when the absolute difference between the query and hit differ greatly. In this instance the I-TASSER generated structure may be represented differently than the query structure.

From this table the user may select desired structures to shuttle forward to TM-align for analysis. The user may also download any of the generated PDBs from this table. When the desired accessions are selected using the check boxes in the left most column of the table, the user may select the “Push to TM-align” button to shuttle the accessions forward to the “TM-align” Selection table.

The user may download all generated protein structures as a Zip file through the “Download Zipped PDB Text Files” button.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov



These Zip Files will contain all generated PDBs from the I-TASSER output in PDF format and will be named by the protein accession.

## *Input Other Protein Structures for TM-align: AlphaFold, RCSB PDB, and Other*

Users may bring in externally generated PDBs with the optional Input Other Protein Structures for TM-align feature. This section consists of a table containing Swiss Prot accessions associated with the Level 1 Full Report output as well as a separate section containing multiple text boxes for submitting PDB information for integration of externally generated structures in TM-align.

Input Other Protein Structures for TM-align: AlphaFold, RCSB PDB, and Other								
Search: Enter keyword								
Swiss Prot Accession	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Scientific Name	Common Name	Protein Name	
<input type="radio"/> A0A2IZYGE1	XP_018875276.1	52137	9595	Mammalia	Gorilla gorilla gorilla	Western lowland gorilla	androgen receptor	
<input type="radio"/> A0A2J8UBB2	XP_009233214.1	140470	9601	Mammalia	Pongo abelli	Sumatran orangutan	androgen receptor isoform X1	
<input type="radio"/> A0A2K5B014	XP_011731144.1	68729	9545	Mammalia	Macaca nemestrina	Pig-tailed macaque	androgen receptor isoform X1	
<input type="radio"/> A0A09RHK8	XP_007990129.1	62302	60711	Mammalia	Chlorocebus sabaeus	Green monkey	androgen receptor	
<input type="radio"/> A0A2K5ZPC3	XP_011835925.1	38457	9568	Mammalia	Mandrillus leucophaeus	Drill	PREDICTED: androgen receptor isoform X1	
<input type="radio"/> A0A2K9I685	XP_011817378.1	38676	336983	Mammalia	Colobus angolensis palliatus	Angolan colobus	PREDICTED: androgen receptor	
<input type="radio"/> A0A2SLJT9	XP_011916275.1	66421	9531	Mammalia	Cercocebus atys	Sooty mangabey	PREDICTED: androgen receptor	
<input type="radio"/> O9Q755	NP_001028083.1	178339	9544	Mammalia	Macaca mulatta	Rhesus monkey	androgen receptor	
<input type="radio"/> A0A2K5EA32	XP_012316261.1	49121	37293	Mammalia	Aotus nancymaeae	Ma's night monkey	androgen receptor isoform X1	
<input type="radio"/> F7GJ39	XP_002762998.1	87664	9483	Mammalia	Callithrix jacchus	White-tufted-ear marmoset	androgen receptor isoform X1	

The user may select the desired Swiss Prot accessions and access the AlphaFold website through the provided link. By selecting a Swiss Prot accession, the table containing text boxes should automatically populate all textboxes except the “PDB” text box.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Input Other Protein Structures for TM-align: AlphaFold, RCSB PDB, and Other

Search: Enter keyword

Swiss Prot Accession	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Scientific Name	Common Name	Protein Name
<input checked="" type="radio"/> A0A2I2YGE1	XP_018875276.1	52137	9595	Mammalia	Gorilla gorilla gorilla	Western lowland gorilla	androgen receptor
<input type="radio"/> A0A2J8UBB2	XP_009233214.1	140470	9501	Mammalia	Pongo abelii	Sumatran orangutan	androgen receptor isoform X1
<input type="radio"/> A0A2K5EG14	XP_011731144.1	68720	9545	Mammalia	Macaca nemestrina	Pig-tailed macaque	androgen receptor isoform X1
<input type="radio"/> A0A0D9RHK8	XP_007990129.1	62302	80711	Mammalia	Chlorocebus sabaeus	Green monkey	androgen receptor
<input type="radio"/> A0A2K5ZPC3	XP_011835925.1	38457	9568	Mammalia	Mandrillus leucophaeus	Drill	PREDICTED: androgen receptor isoform X1
<input type="radio"/> A0A2K5I885	XP_011817378.1	38876	336983	Mammalia	Colobus angolensis palliatus	Angolan colobus	PREDICTED: androgen receptor
<input type="radio"/> A0A2K5LJ79	XP_011916275.1	66421	9531	Mammalia	Cercocebus atys	Sooty mangabey	PREDICTED: androgen receptor
<input type="radio"/> Q5Q795	NP_001028083.1	178339	9544	Mammalia	Macaca mulatta	Rhesus monkey	androgen receptor
<input type="radio"/> A0A2K5EA32	XP_012316281.1	48121	37293	Mammalia	Aotus nancymaae	Ma's night monkey	androgen receptor isoform X1
<input type="radio"/> FTGJ39	XP_002782968.1	57664	9483	Mammalia	Callithrix jacchus	White-tufted-ear marmoset	androgen receptor isoform X1

(1 of 16)

---

Information for TM-align

- Select Source -

Accession:

Protein Name:

Species Tax Id:

Taxonomic Group:

Scientific Name:

Common Name:

PDB:

The ‘PDB’ text box is for the user to copy and paste their externally generated PDB. Note that the PDB should be opened in a txt viewer and the user should select the full text representation of the PDB and past the text in the PDB text box.

## TM-align

TM-align is a protein structure alignment tool that provides scores for each aligned structure on a residue-by-residue basis. TM-align will output a TM-Score (0,1) where 1 is a perfect identical match between the aligned structures.

Users may select accessions in the I-TASSER table after they have completed generating PDB files using the checkboxes and select the ‘Push to TM-align’ button to send the selected PDBs to a queuing table found in the TM-align Selection section of the Level 4 data page before submission. Users may also add externally generated PDBs to the queuing table through the ‘Input Other Protein Structures for TM-align: AlphaFold, RCSB PDB, and Other’ section.

TM-align Selection

Search: Enter keyword

PDB Source	Accession	Restrains Template	Tax ID	Taxonomic Group	Scientific Name	Common Name	Protein Name
<input checked="" type="checkbox"/> I-TASSER	A0A2J8UBB2	-	9501	Mammalia	Pongo abelii	Sumatran orangutan	androgen receptor isoform X1
<input checked="" type="checkbox"/> I-TASSER	KY038911.1	-	8499	Crocodylia	Alligator mississippiensis	American alligator	androgen receptor
<input checked="" type="checkbox"/> I-TASSER	AAC97388.1	-	8105	Amphibia	Xenopus laevis	African clawed frog	androgen receptor alpha isoform
<input checked="" type="checkbox"/> I-TASSER	TFK13170.1	-	85044	Testudinata	Phrynosoma macleodorum	Big-headed turtle	serine/arginine-rich splicing factor 5-like
<input checked="" type="checkbox"/> I-TASSER	NOV50095.1	DAMIAA	12149	Amni	Uta stansburiana	Common muren	ANDR protein
<input checked="" type="checkbox"/> I-TASSER	MEH318736.1	-	7917	Actinopteri	Atractosteus spatula	Alligator gar	ANDR protein
<input checked="" type="checkbox"/> I-TASSER	NOV50095.1	-	12149	Amni	Uta stansburiana	Common muren	ANDR protein
<input checked="" type="checkbox"/> I-TASSER	AA029741.1	-	9594	Mammalia	Homo sapiens	Human	AR protein

(1 of 8)

TM-align Results

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

The user may remove singular PDBs from the queuing table by selecting the checkbox button in the far-left column of the accession's row and selecting the "Remove Selected Row" button. The user may also clear all queued accessions through the "Clear" button.

Once the user has added all the desired PDB structure files to the queuing table, they may select a template accession. This is the structure all other structures will be compared to in TM-align. The user may select any PDB generated by I-TASSER or integrated into the TM-align table from external sources as their template.

Search:	Enter keyword					
<input type="checkbox"/>	AlphaFold	A0A2J8UBB2	-	9001	Mammalia	Pongo abelii
<input type="checkbox"/>	I-TASSER	KYO38361.1	-	8496	Crocodylia	Alligator mississippiensis
<input type="checkbox"/>	I-TASSER	AAC97386.1	-	8355	Amphibia	Xenopus laevis
<input type="checkbox"/>	I-TASSER	TFK15370.1	-	55544	Testudinata	Platysternon megacephalum
<input type="checkbox"/>	I-TASSER			446	Aves	Uria aalge
<input type="checkbox"/>	I-TASSER			17	Actinopteri	Atractosteus spatula
<input type="checkbox"/>	I-TASSER	A0A2J8UBB2(AlphaFold)		46	Aves	Uria aalge
<input type="checkbox"/>	I-TASSER	KYO38361.1 (148:Test_7_28_CL)		96	Mammalia	Homo sapiens
<input type="checkbox"/>	I-TASSER	AAC97386.1 (145:Test_AR_7_13)				

The templates are uniquely named by their accession (I-TASSER SeqAPASS Run Id: User defined name (w/restraint text autogenerated)). Once all desired PDBs are queued and a template chosen, the user may click the "Submit to TM-align" button to submit their PDBs for structural alignment.

### TM-align Results

To view the results for TM-align, select the dropdown menu within the "TM-align Results" table. This dropdown menu will contain all processed TM-align runs pertaining to the current Level 4 report. After selecting the desired run, select the "View TM-align Report" button to bring up the results table.

Tax ID	Taxonomic Group	Susceptibility	Absolute Length
419890	Aves	Y	50.66
135165	Aves	Y	48.06
321034	Aves	Y	49.45
458196	Aves	Y	43.22

The "TM-align Results" table reports the metrics generated by TM-align structural alignment of the chosen PDBs. These results include the L1 value which is the length of the selected PDB and the L2 value, which is the length of the template PDB. The table also reports the I-TASSER metrics reported in the previous "I-TASSER Results" table as well as the TM-align metrics which include "Value 1", "Value 2", and "Average Value". "Value 1" reports the alignment of the template structure to the hit structure, while "Value 2" reports the alignment of the hit structure to the template. The "Average Value" then reports an average of both "Value 1" and "Value 2". Since accessions must go through multiple filters during the Level 1, Level 2, Level 3, and Level 4 analysis, the expectation is that the accessions used for

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

PDB generation are of high quality and relevance. The susceptibility call-out for the “TM-align Results” table is therefore considered “Yes” by default, but experienced users should make their own determination of susceptibility after the alignment of the generated Level 4 structures.

PDB Source	Accession	Average Percent Similarity	Quality	Protein Name	Tax ID	Taxonomic Group	Scientific Name	Common Name	Susceptibility	Absolute Length	Length Cutoff	C-Score	TM-Score	RMSD	Density	L1	L2	Value 1	Value 2	Average Value
1-TASSER	NX000117.1	100.00	Low	ANDR protein	410900	Aves	Saksherpha lutuosus	Song birds	Y	50.05	Check	-3.93	0.29	2.7	0.004	0	0	0.00	0.00	1.00
1-TASSER	NX000293.1	87.34	Low	ANDR protein	150105	Aves	Brachypteryx leucostomus	Shore-lagged ground-spar	Y	48.06	Check	-3.29	0.35	3.3	0.009	491	493	0.90	0.95	0.97
1-TASSER	NX001746.1	92.27	Low	ANDR protein	321054	Aves	Circus hectorius	Bronze-winged wood-sage	Y	49.49	Check	-4.09	0.23	2.9	0.004	491	492	0.97	0.99	0.99
1-TASSER	NX000293.1	84.82	Low	ANDR protein	408100	Aves	Urochloa indica	Mousetails	Y	43.22	Check	-2.77	0.4	3.3	0.016	491	519	0.98	0.91	0.95

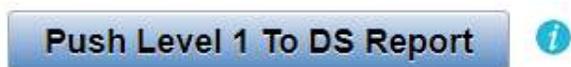
The “TM-align Results” table may be downloaded separately as either an .excel or .csv file through the “Download Table” icons.

### Decision Summary Report

The “Decision Summary (DS) Report” is a feature that gives the user options to design a single output page to concisely view results from all Levels of the SeqAPASS evaluation for completed jobs. The output is customizable to include visualizations and susceptibility predictions that can be downloaded in a PDF format. The “DS Report” page becomes activate when the user utilizes any of the “Push to DS Report” buttons located throughout the toolset. The “DS Report” page will contain a maximum of one Level 1 output (and visualization) and one Level 3 output (and visualization) but can contain multiple Level 2 domain outputs (and their respective visualizations).



To push results from any Level to the DS Report, the user must press the “Push Level # To DS Report” button. The “DS Report” button will become active for the user to view the report settings. The DS Report can be updated as the user changes settings in Level 1, Level 2, and Level 3 (Adding or removing amino acids) but the user must push the updated report to the DS Report again using the “Push Level # To DS Report” button (There will be a notification next to the button if settings have been updated to remind the user to push the report). If the user chooses to change to a different SeqAPASS job (e.g., a different protein accession), the “DS Report” button will become inactive and the user must push the data from the new job to the DS Report as described previously.



#### Level 1 of the Decision Summary Report

Upon clicking the “DS Report” button, the user is brought to a new page that will contain the “Level 1 Report” section of the DS Report which will show all the pertinent information for the query protein and report settings that were pushed to the report. The user can also include the Level 1 visualization in the DS Report by going to the “Level 1 Visualization” page and clicking “Push to Boxplot to DS Report”.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

The default visualization or a user customized visualization will then be inserted in the downloadable DS Report PDF once the radio button is selected.

Once the user is satisfied with the data that has been pushed to the DS Report, the “DS Report” button will bring the user to the “Level 1 Report” section which gives the user customizable options. In the “Level 1 Report” section, there is a series of checkboxes in the “Select Taxonomic Groups (CLASS)” box. Here the user can select which taxonomic group(s) they would like to select and display in the DS Report. Upon selecting the taxonomic group(s), the user can then customize the report in the “Select Species” box, by selecting the checkbox next to the species for which the user would like data from Level 1 displayed in the “Final Decision Summary Report” table at the bottom of the page. The template species will always be selected and cannot be deselected. Specie(s) will be active only when at least one taxonomic group is selected in the “Select Taxonomic Groups (CLASS)” box. Level 1 results for those species selected from the “Select Species” box will be integrated in the “Final Decision Summary Report” table at the bottom of the page (Note: if the user does not push a Level 1 job to the “DS Report” page, there will be no information in that section).

The screenshot displays the 'Level 1 Info' section of the SeqAPASS interface. It features a tabbed interface with 'Level 1 Info' selected. Below the tab is a checkbox labeled 'Add Level 1 Info to Report' which is checked. The main content is divided into two columns: 'Level 1 Query Protein Information' and 'Report Settings'. The 'Query Protein Information' column lists: SeqAPASS ID: 1631, Query Species: Homo sapiens, Query Protein: estrogen receptor isoform 1, Query Accession: NP\_000116.2 (with an 'EXIT' link), Ortholog Count: 410, Protein and Taxonomy Data: 06/08/2020, BLAST Version: 2.10.0, and Software Version: 4.1. The 'Report Settings' column lists: Report Type: Primary, E-Value: 0.01, Sorted By Taxonomic Group: CLASS, Common Domains: 1, Species Read-Across: Y, Cut-off %: 34.43, and Show Only Eukaryotes: Y. Below this is an 'Optional Components' section with a table:

Component	Add to Report
Level 1 Visualization	<input type="checkbox"/>

### Level 1 Info

The Level 1 information section becomes present when either a Level 1 report or a Level 1 visualization is pushed to the DS Report. The information contained in the section includes the “Level 1 Query Protein Information” (i.e., SeqAPASS ID, Query Species, Protein, and Accession, Ortholog Count, Protein and Taxonomy Data, Blast Version and Software Version.) as well as the “Report Settings” (i.e., Report Type, E-Value, Sorted By Taxonomic Group, Common Domains, Species Read-Across, Cut-Off, and Show Only Eukaryotes.) and finally the “Optional Components” section which contains the option to include the “Level 1 Visualization” to the report.

### Including Visualizations in DS Reports

The user can also include the “Level 1 Visualization” by going to the visualization page and either pushing the default visualization or a user modified visualization which will then be attached in the downloaded PDF once the radio button is selected. In the scroll downs, the template species will always be selected and cannot be deselected. Specie(s) will be not active until a taxonomic group box is selected. Once that occurs, those respective species will become active and can be deselected individually or by the select all function. Those species selected will become active in the “Final Decision Summary Report” table at the bottom of the page (Note: if a user pushed only a boxplot to the DS Report, then only the “Level 1 Info” and the “Optional Components” will be active).

**Push Level 1 Boxplot To DS Report**

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Level 1 Report

Select Taxonomic Groups (CLASS)

Select All	Taxonomic Group
<input type="checkbox"/>	Mammalia
<input type="checkbox"/>	Testudines
<input type="checkbox"/>	Aves
<input type="checkbox"/>	Crocodylia
<input type="checkbox"/>	Lepidosauria
<input type="checkbox"/>	Amphibia
<input type="checkbox"/>	Chondrichthyes
<input type="checkbox"/>	Ceratodontimorpha
<input type="checkbox"/>	Coelacanthiformes
<input type="checkbox"/>	Actinopteri
<input type="checkbox"/>	Cladistia
<input type="checkbox"/>	Petromyzontiformes

Select Species

Select All	Species
<input checked="" type="checkbox"/>	Human
<input type="checkbox"/>	Western gorilla
<input type="checkbox"/>	Chimpanzee
<input type="checkbox"/>	Western lowland gorilla
<input type="checkbox"/>	Pygmy chimpanzee
<input type="checkbox"/>	Sumatran orangutan
<input type="checkbox"/>	Bornean orangutan
<input type="checkbox"/>	Rhesus monkey
<input type="checkbox"/>	Sooty mangabey
<input type="checkbox"/>	Crab-eating macaque
<input type="checkbox"/>	Pig-tailed macaque
<input type="checkbox"/>	Ugandan red Colobus

Common Name  
 Scientific Name

Level 1 Info

Add Level 1 Info to Report

---

**Level 1 Query Protein Information**

SeqAPASS ID: 1306

Query Species: Homo sapiens

Query Protein: estrogen receptor Isoform 1

Query Accession: [NP\\_000116.2](#) EXIT

Ortholog Count: 348

Protein and Taxonomy Data: 02/28/2019

BLAST Version: 2.8.1

Software Version: 3.2

**Report Settings**

Report Type: Primary

E-Value: 0.01

Sorted By Taxonomic Group: CLASS

Common Domains: 1

Species Read-Across: Y

Cut-off %:33.93

Show Only Eukaryotes: Y

Optional Components

Component	Add to Report
Level 1 Visualization	<input type="checkbox"/>

## Level 2 of DS Report

The Level 2 section of the DS Report contains all the domains that have been pushed to the report. There can be multiple domains present in the section once they have been run and pushed individually to the report. The user can also include each respective “Level 2 Visualization” by going to the visualization page and either pushing the default visualization or a user modified visualization which will then have the option to be attached in the downloaded PDF. Once a domain is selected, it will appear in the “Final Decision Summary Report” table at the bottom of the page (Notes: if the user does not push a Level 2 run to the DS Report page, there will be no information in that section. If a visualization is pushed to the DS Report before a Level 2 report, the domain will be present along with the “Add Visualization to Report” button being active.).

Level 2 Report

Select Level 2 Domains

Add to Final Decision Report	Domain	Optional Components	
Select All		Add Info to Report	Add Visualization to Report
<input checked="" type="checkbox"/>	(316) cd06931, NR_LBD_HNF4_Ilike, The ligand binding domain of hepatocyte nuclear factor 4, which is explosively expanded in nematodi	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	(310) cd06949, NR_LBD_ER, Ligand binding domain of Estrogen receptor, which are activated by the hormone 17beta-estradiol (estroge	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Level 3 of DS Report

The Level 3 section of the DS Report contains all the information for the query protein and report settings that were pushed to the report. It also contains the amino acids that were updated in the report and pushed over. New amino acids will need to be pushed over to the DS Report. The Yes (Y) or No (N) susceptibility will be displayed in the “Final Decision Summary Report” table. The user can also include the “Level 3 Visualization” by going to the visualization page pushing a user modified visualization which will then have the option to be attached in the downloaded PDF. (Notes: if the user does not push a Level 3 run to the DS Report page, there will be no information in that section. Also, if a “Level 3 Visualization” (Heat Map) is pushed before a Level 3 report, the “Level 3 Info” will be populated with that respective run’s information.)

Component	Add to Report
Level 3 Visualization	<input type="checkbox"/>

### Final Decision Summary Report Table

The “Final Decision Summary Report” table contains the important data and susceptibility predictions for each level run, for all the species selected in the Level one section. The table takes the susceptibility prediction for each run and easily displays the results for a quick interpretation. The complete table can be either saved as an excel spreadsheet or .csv file. It will also be added into the PDF when downloaded.

Each selected specie(s) will have its own respective row which contains the information that has been pushed to the “Final Decision Summary Report” table. The columns will show the Data Version, NCBI Accession, Filtered Taxonomic Group, Species, Protein Name, Level 1 Susceptibility Prediction as Yes (Y) or No (N), Level 2 Common Domain(s) Name and respective Susceptibility Prediction as Yes (Y) or No (N), Level 3 Template Species, and Level 3 Amino Acid Susceptibility Prediction as Yes (Y) or No (N). (A few things to note: if there are multiple domains pushed to the "Final Decision Summary Report" table, each domain will have their own column. Also, for species to have either a Yes (Y) or No (N) susceptibility prediction in the table, they must be pushed to the report from the Level 3 run as well as selected in the Level 1 taxonomic groups/species selection. If a species was not included in the Level 3 report that was pushed but is included in the “Final Decision Summary Report” table, they will receive a NA for their Level 3 susceptibility prediction.)

Data Version	NCBI Accession	Filtered Taxonomic Group	Species	Protein	Susceptible (Y/N)	(345) cd06157, NR_LBD, The ligand binding domain of nuclear receptors, a family of ligand-activated transcription regulators	Level 3 Template	Level 3 Amino Acids (Y/N)
5	NP_000116.2	Mammalia	Human	estrogen receptor isoform 1	Y	Y	Homo sapiens	Y

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Download DS Report as PDF

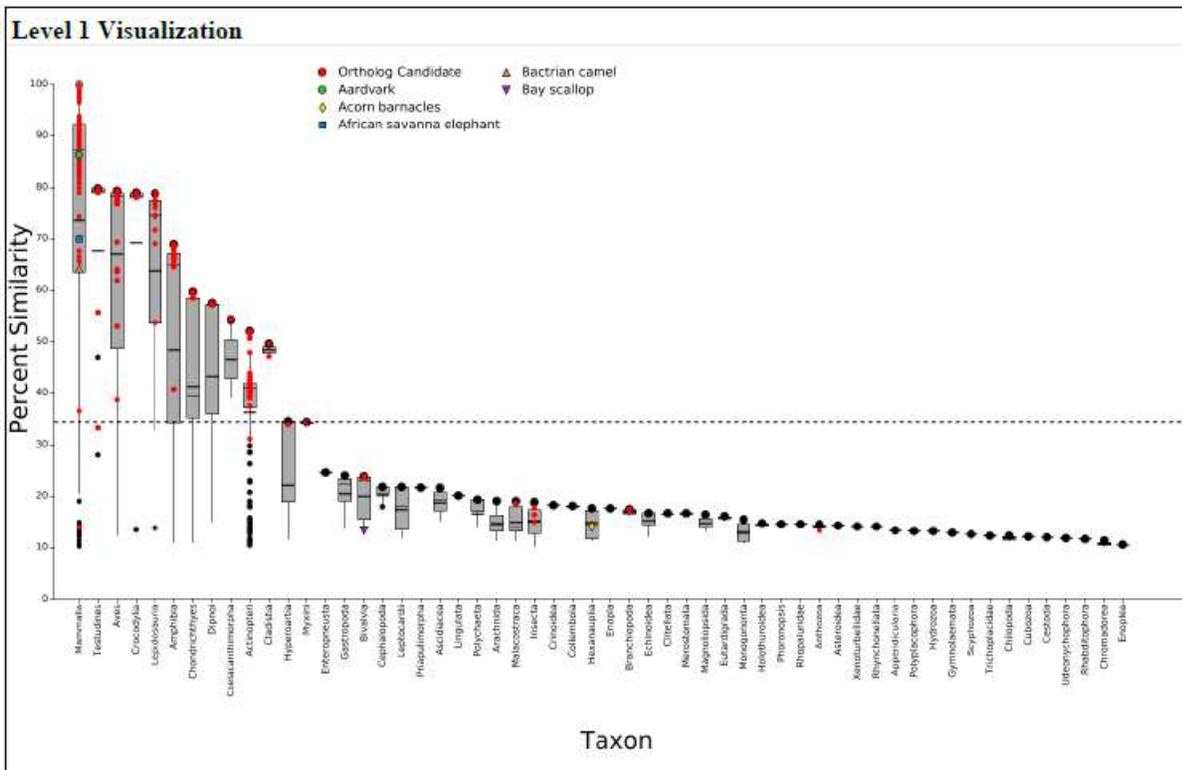
To capture all the data pushed to the DS Report as a PDF, press the “Download DS Report” button. The DS Report PDF will match the data on the DS Report page and will include the visualizations if selected by the user. The information for each Level that is pushed to the downloaded DS Report PDF include all the Query Protein Information for that respective protein, domain(s), and template protein. (Note: Once the PDF is created and the DS Report page has been updated, the user must redownload the PDF to have the most up to date version of the page.)

### Level 1 of DS Report PDF

The Level 1 section of the DS Report PDF will contain all the “Level 1 Query Protein Information” along with the Level 1 “Report Settings” for that respective protein’s run. This information will not be present if no Level 1 run information or Level 1 visualization is pushed to the DS Report PDF.

### Level 1

<p><b><u>Level 1 Query Protein Information</u></b></p> <p>SeqAPASS ID: 1679</p> <p>Query Species: Homo sapiens</p> <p>Query Protein: estrogen receptor isoform 1</p> <p>Query Accession: NP_000116.2</p> <p>Ortholog Count: 410</p> <p>Protein and Taxonomy Data: 06/08/2020</p> <p>BLAST Version: 2.10.0</p> <p>Software Version: 4.1</p>	<p><b><u>Report Settings</u></b></p> <p>Report Type: Primary</p> <p>E-value: 0.01</p> <p>Sorted By Taxonomic Group: CLASS</p> <p>Common Domains: 1</p> <p>Species Read-Across: Y</p> <p>Cut-off %: 34.43</p> <p>Show Only Eukaryotes: Y</p>
--	---



## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Level 2 of DS Report PDF

The Level 2 section of the DS Report PDF will contain all the “Level 2 Query Protein Information” along with the Level 2 “Report Settings” for that respective domain’s run. The Level 2 information will not be present if no Level 2 run information or Level 2 visualization is pushed to the DS Report PDF. Each domain can have its own respective visualization that can be added to the DS Report PDF by selecting the “Add Visualization to Report” button in the DS Report page.

Level 2	
<b><u>Level 2 Query Protein Information</u></b> SeqAPASS ID: 1653 Query Species: Homo sapiens Query Domain: (345) cd06157, NR_LBD, The ligand binding domain of nuclear receptors, a family of ligand-activated transcription regulators Query Accession: NP_000116.2 Ortholog Count: 410 Protein and Taxonomy Data: 06/08/2020 BLAST Version: 2.10.0 Software Version: 4.1	<b><u>Report Settings</u></b> Report Type: Primary E-value: 10.0 Sorted By Taxonomic Group: CLASS Species Read-Across: Y Cut-off %: 55.00 Show Only Eukaryotes: Y

Level 2	
<b><u>Level 2 Query Protein Information</u></b> SeqAPASS ID: 1653 Query Species: Homo sapiens Query Domain: (341) cd06929, NR_LBD_F1, Ligand-binding domain of nuclear receptor family 1 Query Accession: NP_000116.2 Ortholog Count: 409 Protein and Taxonomy Data: 06/08/2020 BLAST Version: 2.10.0 Software Version: 4.1	<b><u>Report Settings</u></b> Report Type: Primary E-value: 10.0 Sorted By Taxonomic Group: CLASS Species Read-Across: Y Cut-off %: 42.03 Show Only Eukaryotes: Y

### Level 3 of DS Report PDF

The Level 3 section of the DS Report PDF will contain all the “Level 3 Template Protein Information” along with the Level 3 “Selected Amino Acids” for that respective run. This information will not be present if no Level 3 run information or Level 3 visualization is pushed to the DS Report PDF. The run can have a visualization “Heat Map” that can be added to the DS Report PDF by selecting the “Add Visualization to Report” radio button.

Level 3	
<b><u>Selected Amino Acids</u></b> 5L, 57G, 120F, 177E	<b><u>Level 3 Template Protein Information</u></b> SeqAPASS ID: 1653 Template Species: Homo sapiens Template Protein: [NP_000116.2] estrogen receptor isoform 1 Protein and Taxonomy Data: 06/08/2020 BLAST Version: 2.10.0 Software Version: 4.1

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

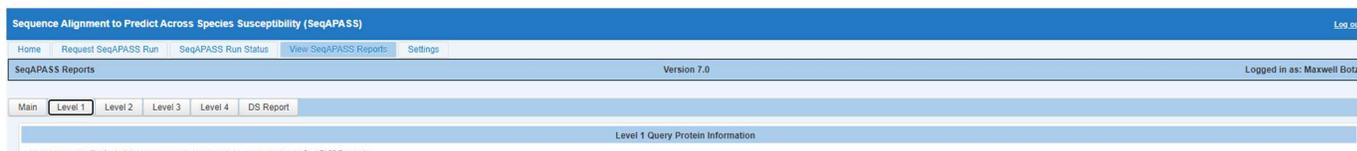
*Final DS Report Table in DS Report PDF*

The Final Decision Summary Report table will display the species that were selected for the Level 1 section of the DS report. It can display the specie’s respective “Protein”, “Level 1 Susceptibility (Y/N)”, common domain(s), “Level 3 Template”, and “Level 3 Susceptibility” all depending on what is selected from the DS Report set up.

Final Decision Summary Report			
Species	Protein	Level 1 Susceptible (Y/N)	(345) cd06157, NR_LBD, The ligand binding domain of nuclear receptors, a family of ligand-activated transcription regulators
Human	estrogen receptor isoform 1	Y	Y
Western gorilla	estrogen receptor alpha	Y	Y
Chimpanzee	estrogen receptor isoform X2	Y	Y
Western lowland gorilla	estrogen receptor isoform X2	Y	Y
Pygmy chimpanzee	estrogen receptor isoform X2	Y	Y
Bornean orangutan	estrogen receptor alpha	Y	Y
Sumatran orangutan	estrogen receptor isoform X2	Y	Y
Sooty mangabey	PREDICTED: estrogen receptor isoform X2	Y	Y
Rhesus monkey	estrogen receptor isoform X2	Y	Y

### Moving Between Level 1, Level 2, Level 3, Level 4, and Decision Report Data Pages

As a user chooses to view Level 1, Level 2, Level 3, or Level 4 data in the “View SeqAPASS Reports” tab, new buttons become available for allowing the user to move between Levels of an analysis. The Decision Report data page will become active once a user pushes a finished run using the “Push Level # To DS Report” button.



The user can use the “Main” button to return to the list of completed Level 1 runs and select a different query accession to view. The “Level 1” button brings the user to the Level 1 data page, where the user can set up queries for Level 2, Level 3, and if they have access, Level 4, as well as select the button to view Level 2, Level 3, and Level 4 data pages. Open Level 1, Level 2, Level 3, and Level 4 pages remain open until the user selects a different run to view on the “Main” page. Moving between tabs, such as “Home,” Request SeqAPASS Run,” and “SeqAPASS Run Status”, does not close the Level 1, Level 2, Level 3, or Level 4 pages that have been opened.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

**Note:** If the user logs out of the SeqAPASS tool, upon logging back in, the data will reset to default settings. Therefore, the View SeqAPASS Reports tab will not display the “Main,” “Level 1,” “Level 2,” “Level 3,” or “Level 4” buttons, until a query is chosen and Level 2, Level 3, and/or Level 4 pages are opened.

## Search, View, and Download Data Tables

The user can use the “Search” box to enter text to search the table. Further, the user can use the arrow buttons and page numbers on the bottom of the screen to view all data and the drop-down to expand the table to 10, 20, or 50 rows. There are also left and right scroll bars at the bottom of the tables to allow the user to view all columns of the table.

Search using text box on top of tables:



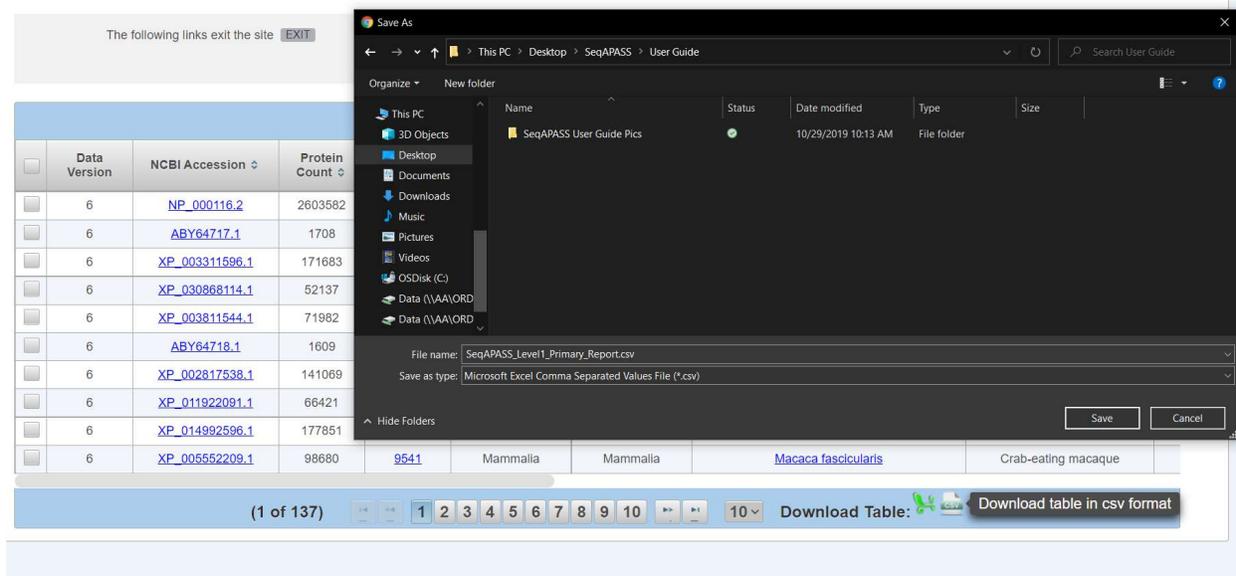
Options for viewing data:



All data tables in the SeqAPASS tool can be downloaded as Excel or csv files. The icons for downloading the files are present on the bottom right-hand side of all tables. Click the icon to download data.



Upon selecting a csv file, the user can choose to save or open the file. Each file is appropriately named by Level of the SeqAPASS evaluation and report type.



## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Upon selecting a .xls file, the user can save the report to their desired location. Each file is appropriately named by Level of the SeqAPASS evaluation and report type.

The screenshot displays the SeqAPASS web application interface. At the top, there are navigation tabs: 'Full Report', 'Susceptible = Y, Ortholog Count = 1', and 'Show Only Eukaryotes'. The main content area is titled 'Level 2 Data - Primary' and contains a search bar with the text 'Search: Enter keyword'. Below the search bar is a table with the following columns: 'Data Version', 'NCBI Accession', 'Protein Count', 'Species Tax ID', 'Taxonomic Group', and 'Filtered Taxonomic Group'. The table contains 10 rows of data, all with a 'Data Version' of 4 and a 'Taxonomic Group' of 'Mammalia'. A 'Download Table' button is located at the bottom right of the table, with a tooltip that reads 'Download table in excel(.xls) format'. A 'Save As' dialog box is open over the table, showing the file name 'SeqAPASS\_Level2\_Primary\_Report.xls' and the save type 'Microsoft Excel 97-2003 Worksheet (\*.xls)'. The dialog box also shows the current directory as 'Desktop > SeqAPASS Reports'.

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group
4	<a href="#">NP_000116.2</a>	1265506	<a href="#">9606</a>	Mammalia	Mammalia
4	<a href="#">XP_014992596.1</a>	88400	<a href="#">9544</a>	Mammalia	Mammalia
4	<a href="#">ABY64721.1</a>	931	<a href="#">9534</a>	Mammalia	Mammalia
4	<a href="#">XP_003255939.1</a>	38964	<a href="#">61853</a>	Mammalia	Mammalia
4	<a href="#">XP_025240309.1</a>	52618	<a href="#">9565</a>	Mammalia	Mammalia
4	<a href="#">XP_003811544.1</a>	51891	<a href="#">9597</a>	Mammalia	Mammalia
4	<a href="#">XP_011922091.1</a>	66748	<a href="#">9531</a>	Mammalia	Mammalia
4	<a href="#">ABY64717.1</a>	2023	<a href="#">9593</a>	Mammalia	Mammalia
4	<a href="#">XP_002817538.1</a>	145798	<a href="#">9601</a>	Mammalia	Mammalia
4	<a href="#">XP_011852180.1</a>	38580	<a href="#">9568</a>	Mammalia	Mammalia

### Log out

The user can log out from any page in SeqAPASS, by clicking the “Log out” link on the upper right-hand side of the page. If a user clicks Log out and then Logs back in, all settings will be set back to default. User can log out at any time by clicking the “[Log out](#)” link on the upper right-hand side. Any successfully submitted queries that were requested prior to logging out will continue running and when completed, will be available to the user in the “View SeqAPASS Reports” tab.

The screenshot shows the SeqAPASS navigation bar. It includes the following elements: 'Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)' on the left, 'Log out' on the right, and a set of navigation tabs: 'Home', 'Request SeqAPASS Run', 'SeqAPASS Run Status', 'View SeqAPASS Reports', and 'Settings'. Below the tabs, it displays 'SeqAPASS Reports', 'Version 7.0', and 'Logged in as: Maxwell Botz'.

### Pop-up Messages

The Spinning Wheel pop-up is used as an indicator to alert the user that an action is taking place, where the interface of the SeqAPASS tool is contacting the backend database. For example, upon clicking the “SeqAPASS Run Status” tab, “Refresh Data” button, “View Level 2 Data” button, or “View Level 3 Data” button the Spinning Wheel will pop-up and disappear from the screen. There are multiple other instances where the spinning wheel is used as an indicator to the user that an action is occurring.



Pop-up messages are meant to guide the user to submit the correct information for a query, inform the user of a successful or failed query submission, or otherwise inform the user of an error. All pop-up messages will appear for 10 seconds on the upper right-hand side of the screen, and then disappear. If the

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

user would like to close the message before the 10 seconds is up, click on the message and an “x” will appear of the upper right-hand corner of the message box. Click the x to close the message.

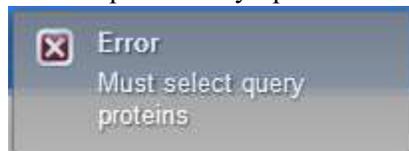
In the “Request SeqAPASS Run” tab, Compare Primary Amino Acid Sequences “By Species” page, a successful Level 1 query submission will display a pop-up message indicating that the query has been submitted to the run queue or if “existing” message appears indicating that the accession has been ran previously either by a user and is available to view.



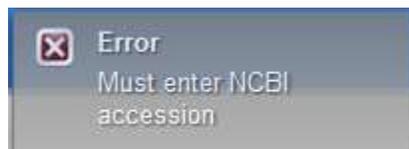
OR



User did not select any query proteins from the “Request SeqAPASS Run” tab, Compare Primary Amino Acid Sequences “By Species” or “By Accession” page, and clicked “Request Run” button.



OR



If the user enters non-sense text (or any text that is not an NCBI accession) into the “NCBI Protein Accession” text box for submitting a Level 1 query in the “Request SeqAPASS Run” tab, in the Compare Primary Amino Acid Sequences “By Accession” page, and clicked “Request Run” button, the message below will pop-up indicating that the Accession entered is not in the SeqAPASS database.



In the “View SeqAPASS Reports” tab, Level 1 page, if a user clicks “View Level 2 Data,” a successful Level 2 query submission will display a pop-up message indicating that the query has entered the run queue.



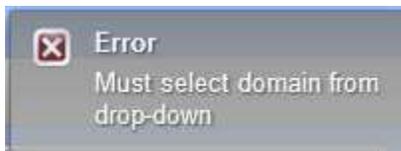
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: [LaLone.Carlie@epa.gov](mailto:LaLone.Carlie@epa.gov)

In the “View SeqAPASS Reports” tab, Level 1 page, if a user selects a domain that has already been submitted (but not completed) and clicks “Request Domain Run” a message for successful Level 2 query submission will display a pop-up message indicating that the query has entered the run queue



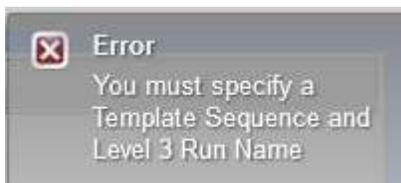
In the “View SeqAPASS Reports” tab, Level 1 page, if a user clicks “View Level 2 Data” without selecting a domain to view from the drop-down, the message below will pop-up to indicate that the user must select a domain.



In the “View SeqAPASS Reports” tab, Level 1 page, a successful Level 3 query submission will display a pop-up message indicating that the query has entered the run queue.



In the “View SeqAPASS Reports” tab, Level 1 page, if a user fails to type a user defined Level 3 Run Name, the message below will pop-up to indicate that the user must do so.



In the “View SeqAPASS Reports” tab, Level 1 page, if a user fails to select species from the Level 1 Data table to be compared with the template sequence, the message below will pop-up.



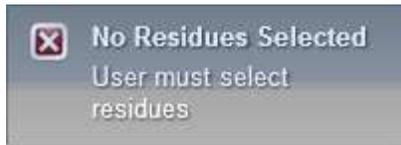
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: [LaLone.Carlie@epa.gov](mailto:LaLone.Carlie@epa.gov)

In the “View SeqAPASS Reports” tab, Level 1 page, if a user fails to select a Level 3 Run Name from the Choose Query to View drop-down and clicks the “View Level 3 Date” button, the message below will pop-up.



In the “View SeqAPASS Reports” tab, “Level 3 Template Protein Information” data page, if a user fails to select amino acid residues using the “Select Amino Acid Residues” shuttle and clicks the “View Level 3 Date” button, the message below will pop-up.



## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Documentation

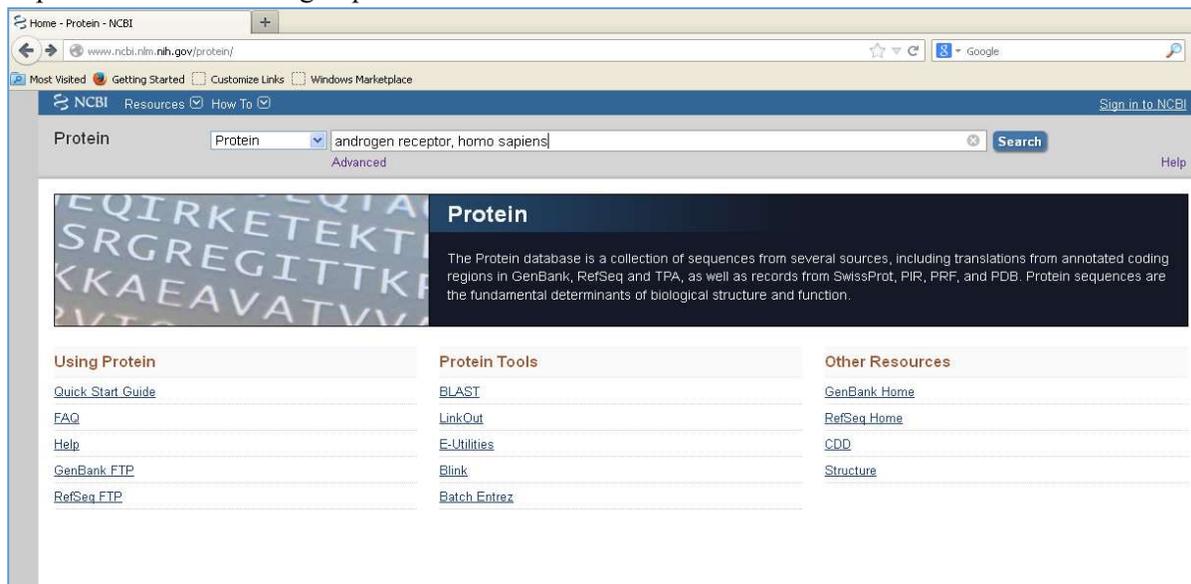
**Query Species:** The selection of the query species for a SeqAPASS analysis is dependent upon the question the user is addressing. For example, the query species can be the target species (i.e., human or companion animal in the case of drugs; or insect, plant, fungus, or pest in the case of pesticides) or, depending on the application of the susceptibility prediction, the query species may be a species known or hypothesized to be sensitive to a chemical acting on the protein molecular target of interest. There may be instances where a protein for the species of interest has not been sequenced, in this case it may serve the users purpose to identify another taxonomically related species from the same organism Class, Order, Family, or Genus as a surrogate query species. In certain cases, when there is interest in the susceptibility of a particular species (e.g., honey bee) and in the case that there are numerous potential target species (e.g., neonicotinoids are intended to cause mortality in a number of pest insects) the species of particular concern may serve as the query species.

**Query Protein:** SeqAPASS can be queried with any protein sequence available in the NCBI protein GenBank database, by protein name, or NCBI Accession. It is suggested that the user of SeqAPASS examines their query protein and species in the NCBI protein database prior to submitting a run to SeqAPASS (use NCBI link on query page). It is not uncommon for a protein of a specific species to be represented by more than one sequence. In such cases there are some guiding principles for identification of the best sequence available for the SeqAPASS run.

**General guidelines:** These guidelines describe best practices for identifying the most useful sequence for a species susceptibility prediction in SeqAPASS, however, in some cases, limited sequence information is available and therefore less desirable sequences may be used. It is up to the user of SeqAPASS to recognize the quality and limitations of the sequence chosen for the SeqAPASS query. The information about a particular protein can be found on the Protein page in the NCBI database

(<http://www.ncbi.nlm.nih.gov/protein/>).

<http://www.ncbi.nlm.nih.gov/protein/>



Search for a protein of interest using protein name and/or species of interest: For the example above, multiple hit proteins were identified.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

The screenshot shows the NCBI Protein search interface. The search term is 'androgen receptor, homo sapiens'. The results list five entries, each with a checkbox, accession number, and length. The first entry is 'Full=Nuclear receptor subfamily 3 group C member 4' (919 aa protein). The second entry is 'androgen receptor [Homo sapiens]' (917 aa protein). The third entry is 'androgen receptor, partial [Homo sapiens]' (2 aa protein). The fourth entry is 'androgen receptor [Homo sapiens]' (906 aa protein). The fifth entry is 'androgen receptor [Homo sapiens]' (917 aa protein). The interface includes filters for species, enzyme types, source databases, sequence length, molecular weight, release date, and revision date. There are also options to 'Find related data' and 'Search details'.

Select one of the proteins by clicking on the link shown above to see detailed information about the protein

The screenshot shows the detailed NCBI protein page for 'androgen receptor [Homo sapiens]' (GenBank: AAA51771.1). The page includes a 'FASTA' link and a 'Go to' section. The main content is organized into sections: LOCUS, DEFINITION, ACCESSION, VERSION, DBSOURCE, KEYWORDS, SOURCE, ORGANISM, REFERENCE, and FEATURES. The DEFINITION section states: 'androgen receptor [Homo sapiens]'. The REFERENCE section lists two references, with the first being 'Tilley, W.D., Marcelli, M., Wilson, J.D. and McPhaul, M.J. Characterization and expression of a cDNA encoding the human androgen receptor. Proc. Natl. Acad. Sci. U.S.A. 86 (1), 327-331 (1989)'. The FEATURES section includes 'source' and 'location/Qualifiers'. On the right side, there are sections for 'Change region shown', 'Customize view', 'Analyze this sequence' (with options for Run BLAST, Identify Conserved Domains, Highlight Sequence Features, and Find in this Sequence), 'Protein 3D Structure' (with a 3D model image and details like 'Targeting The Binding Function 3 (h5) Site Of The Human Androgen Receptor PDB: 4HLW'), 'Articles about the AR gene' (with a list of articles), and 'Identical proteins for AAA51771.1'.

**Guiding principles:** On the NCBI protein page, rows to examine include: “DEFINITION,” “REFERENCES,” COMMENTS,” and “FEATURES.” The information provided in these rows can aid a SeqAPASS user in the identification of an ideal query sequence for SeqAPASS.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

It is desirable to:

- a. Use accessions with the following prefix: NP\_
- b. Avoid use of protein sequences labeled “partial,” “PREDICTED,” “PROVISIONAL,” “INFERRED,” or “hypothetical”
- c. Avoid using those labeled “TPA” (Third Party Annotation), however if TPA is all that is available “TPA: experimental” would be preferred over “TPA: inferential”
- d. Look at the date associated with the protein in the “LOCUS” row of the detailed protein page. A more recent date can have the most up-to-date annotation of the protein. Under the “DBSOURCE” row of the detailed protein page other accessions associated with past protein sequences can be viewed. Many times, if the “xrefs” row is heavily populated and has the most recent annotation update date, it is likely to be the best sequence to use as a query sequence in SeqAPASS.
- d. Short sequences should be avoided when possible as query sequences. Many times, if one selects the protein from the protein output derived from the NCBI protein database query, they will find that the short sequence is actually a partial sequence described in the “DEFINITION” row of the Protein page.
- e. Unless there is reason for doing so (based on the question the user is trying to address), splice-variants labeled in “FEATURES” rows of the Protein page as “alternatively spliced” would be less desirable
- f. It is important to check the references associated with the selected query protein. In some cases, certain sequences are associated with sensitivity to a given chemical. This can be particularly useful when predicting susceptibility to pesticides, where certain strains of insects are produced to be readily sensitive or insensitive to a chemical.
- g. A secondary check of the sequence used in the SeqAPASS run would be to look at the output derived and see whether ortholog candidates were detected. Ideally a preferential sequence would have more ortholog candidates identified.

**Important Note:** To identify which query protein has the greatest number of Ortholog Candidates the user can choose to submit multiple proteins with the same species and protein. Upon the Level 1 runs completing for those similar proteins, the user can then select the “View SeqAPASS Reports” tab and look at the table for “Ortholog Count” the protein with the highest number is likely to be the most appropriate query species for a SeqAPASS evaluation.

Example: Androgen receptor, *Homo sapiens*

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

NCBI Resources How To Sign in to NCBI

Protein Protein Search Advanced Help

Display Settings: GenPept Send to:

## androgen receptor [Homo sapiens]

GenBank: [AAA61771.1](#)  
 FASTA Graphics

Go to:

LOCUS AAA61771 917 aa linear PRI 21-OCT-1994

**DEFINITION** androgen receptor [Homo sapiens].

**ACCESSION** AAA61771

**VERSION** AAA61771.1 GI:179872

**RESOURCE** locus: HUMAN accession: M21748.1

**KEYWORDS** .

**SOURCE** Homo sapiens (human)

**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Heterochin; Homidae; Homo.

**REFERENCE** 1 (residues 1 to 917)

**AUTHORS** Tilley, B.D., Marcelli, M., Wilson, J.D., and McPhaul, M.J.

**TITLE** Characterization and expression of a cDNA encoding the human androgen receptor

**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 86 (1), 927-931 (1989)

**PMID** 2511578

**REFERENCE** 2 (sites)

**AUTHORS** Marcelli, M., Tilley, B.D., Wilson, J.D., Griffin, J.E., Wilson, J.D., and McPhaul, M.J.

**TITLE** Definition of the human androgen receptor gene structure permits the identification of mutations that cause androgen resistance: premature termination of the receptor protein at amino acid residue 588 causes complete androgen resistance

**JOURNAL** Mol. Endocrinol. 4 (6), 1105-1116 (1990)

**PMID** 2293028

**COMMENT** [8] sites; androgen resistant mutation. Draft entry and computer-readable sequence [1] kindly submitted by M.J. McPhaul, 09-DEC-1988.

**FEATURES**

source

Location/Qualifiers

1..917

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="Xq11.2-q12"

/sex="male"

/biotype="prostate"

1..917

/product="androgen receptor"

6..446

/region\_name="Androgen\_recep"

/note="Androgen receptor; p140R88"

/db\_xref="CID:111097"

552..839

/region\_name="MR\_TD\_ARD"

/note="DNA-binding domain of androgen receptor (AR) is composed of two C4-type zinc fingers: c407L79"

/db\_xref="CID:142547"

order(557,560,574,577,592,599,609,612)

/site\_type="other"

/note="zinc binding site [ion binding]"

/db\_xref="CID:142547"

order(566..569,576,578..579,582..583,591,606..607,610,612)

/site\_type="DNA binding"

/note="DNA binding site [nucleotide binding]"

/db\_xref="CID:142547"

order(592..596,598..600,605,608)

/site\_type="other"

Change region shown

Customize view

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Protein 3D Structure

Targeting The Binding Function 3 (b13) Site Of The Human Androgen PDB: 4HLW Source: Homo sapiens Method: X-Ray Diffraction Resolution: 2.5 Å

See all 64 structures...

Articles about the AR gene

Repression of cell proliferation and androgen receptor activity in pr [Anticancer Res. 2013]

TALEN-engineered AR gene rearrangements reveal endo [Proc Natl Acad Sci U S A. 2013]

Androgen receptor (AR) positive vs negative roles in prostate car [Cancer Treat Rev. 2014]

See all...

Identical proteins for AAA61771.1

androgen receptor [Homo sapiens] [AAA61772]

See all...

Pathways for the AR gene

Integrated Breast Cancer Pathway

SIDS Susceptibility Pathways

Nuclear Receptors

See all...

Reference sequence information

RefSeq genomic sequence

See the genomic reference sequence for the AR gene (NG\_009014.2).

RefSeq protein isoforms

See 4 reference sequence protein isoforms for the AR gene.

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Example cont:

```

/!note="dimer interface [polypeptide binding]"
/db_xref="CID:142547"
670..915
/region_name="MR_LED_AR"
/!note="Ligand binding domain of the nuclear receptor
androgen receptor, ligand activated transcription
regulator: c070792"
/db_xref="CID:122758"
order(699,702..703,705..706,709,729..740,749..744,747,750,
762,778,785,871,875)
/site_type="other"
/!note="Ligand binding site [chemical binding]"
/db_xref="CID:122758"
order(711,714,718,724,728,722,736,891..892,895..896)
/site_type="other"
/!note="coactivator recognition site [polypeptide binding]"
/db_xref="CID:122758"
1..317
/!gene="AR"
/!code_by="M21748.1.163..1916"
/db_xref="GDB:600-120-556"
ORIGIN
1 mevqlglgrv yppppoktyr gafgnlgrv reviqppgr hpeaaasapp ga3llllqqg
61 qgggggggggg qgggggggts prgggggggs dgggahrrg ptgyvlvdes ggsapgsal
121 tchpergcvp epgaaavaak glpqqipapp deddaapst llllgttpp lscsadhkd
181 llraaromgl lgggggavss egsrgare rrgpsskd nylgpcstl dshelkav
241 svmglgvsa lehlrpgql rgsmpgall gpppavpqp caplactqs llldtagkt
281 edaaryptik ggytqlgrv slgcrpsaa gsgtlelps tlllyksqal deaaygrd
361 yynfpialag pppppppppp hariklerpl dgsaaaaa acryydlas lhgavagpg
421 sgrpsaaas rshelitate gglyspcggg gggggggggg gggggggggg
481 trppgglagg esdtepdw ypgmvsrv yprtcvks mggmndryg pygdwrlta
541 edvlpdyy fppgtcllc gdeasgchy atcgstovt hraaegkqk ylcasmdec
601 sdrfmlmp scrlfeyea gnlgarlk klgnlkgrv gaaatpvt eeteqhtov
661 hiegrtgp flrvltaep qvcaghdw qpdiaall: lnelgral vbovmakal
721 pgrmlhvdd gnavigyom glvntamgr stromuml yfapdlvne ymbkrmys
781 qcvmhrlqg efgolqtpp eflmkalll fziipvdgk ngtfdelm ryikeldrii
841 ackrlmptc rryfqltkl ldsqpiare lhgfdlll khmvsvdp emmaeivsg
901 vpkilgkqv pyfheq
//

```

More about the AR gene

The androgen receptor gene is more than 90 kb long and codes for a protein that has 3 major functional domains: the N-terminal domain, DNA-binding domain, and the LBD. Also Known As: RP11-383C12.1, AIS, DHT...

---

Homologs of the AR gene

The AR gene is conserved in Rhesus monkey, dog, cow, mouse, rat, and chicken.

---

LinkOut to external resources

A selection of literature about the proteins [GoPubMed Proteins]

Transcript/Protein Information [PANTHER Classification System]

Transcript/Protein Information [PANTHER Classification System]

biochemicals [ExactAntigen/Labome]

antibody review [ExactAntigen/Labome]

others [ExactAntigen/Labome]

antibody [ExactAntigen/Labome]

cDNA clone [ExactAntigen/Labome]

protein and peptide [ExactAntigen/Labome]

ELISA and assay kit [ExactAntigen/Labome]

h. If multiple proteins appear to be the best query protein for SeqAPASS, the sequences can be aligned using NCBI's COBALT. Enter (copy and paste from NCBI protein search list) accessions and align.

**COBALT**
Constraint-based Multiple Alignment Tool
My NCBI  
[Sign In] [Help]

Cobalt Constraint-based Multiple Protein Alignment Tool

COBALT computes a multiple protein sequence alignment using conserved domain and local sequence similarity information. [Reset page](#)

Enter Query Sequences

Enter at least 2 protein accessions, gis, or FASTA sequences Clear

P10275.2  
AAA51772.1  
AAA51780.1  
AAA51771.1  
AAA51729.1  
AAD45921.1  
AAA51886.1

Or, upload FASTA file  No file selected.

Job Title

Show results in a new window

[Advanced parameters](#)

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Alignment page will be generated

**COBALT** Constraint-based Multiple Alignment Tool

Home Recent Results Help

Phylogenetic Tree Edit and Resubmit Download

- Cobalt RID EMV7SF1X211 (7 seqs)

All queries form only one cluster. No domain information was used for generating constraints. Decreasing maximum in-cluster distance or turning off query clustering option may improve results.

Descriptions Select All Re-align Alignment parameters

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Accession	Description	Links
<input checked="" type="checkbox"/> P10275.2	RecName: Full=Androgen receptor; AltName: Full=Dihydrotestosterone receptor; AltName: Full=Nucle	<a href="#">GM</a>
<input checked="" type="checkbox"/> AAA51772.1	androgen receptor [Homo sapiens] >gb AAA51771.1  androgen receptor [Homo sapiens]	<a href="#">G</a>
<input checked="" type="checkbox"/> AAA51780.1	androgen-receptor [Homo sapiens]	<a href="#">GM</a>
<input checked="" type="checkbox"/> AAA51771.1	androgen receptor [Homo sapiens] >gb AAA51772.1  androgen receptor [Homo sapiens]	<a href="#">GM</a>
<input checked="" type="checkbox"/> AAA51729.1	androgen receptor [Homo sapiens] >sp P10275.2 ANDR_HUMAN RecName: Full=Androgen receptor;	<a href="#">GM</a> <a href="#">PubChem BioAssay Info linked to AAA51729.1</a>
<input checked="" type="checkbox"/> AAD45921.1	androgen receptor [Homo sapiens]	<a href="#">GM</a>
<input checked="" type="checkbox"/> AAA51886.1	androgen receptor [Homo sapiens] >sp P10275.2 ANDR_HUMAN RecName: Full=Androgen receptor;	<a href="#">G</a> <a href="#">PubChem BioAssay Info linked to AAA51886.1</a>

Alignments Select All Re-align Mouse over the sequence identifier for sequence title

View Format: Compact Conservation Setting: 2 Bits

<input checked="" type="checkbox"/> P10275	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000qET	80
<input checked="" type="checkbox"/> AAA51772	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000-ET	79
<input checked="" type="checkbox"/> AAA51780	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000----ET	75
<input checked="" type="checkbox"/> AAA51771	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000-ET	79
<input checked="" type="checkbox"/> AAA51729	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000qET	80
<input checked="" type="checkbox"/> AAD45921	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000----ET	75
<input checked="" type="checkbox"/> AAA51886	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000qET	80
<input checked="" type="checkbox"/> P10275	81	SPRQ0000QGEDGSPQAHRRGPTGYLVLDEEQPSPQSALECHPERGCVPEPGAAVAASKGLPQQLPAPPEDEDSAAFS	160
<input checked="" type="checkbox"/> AAA51772	80	SPRQ0000QGEDGSPQAHRRGPTGYLVLDEEQPSPQSALECHPERGCVPEPGAAVAASKGLPQQLPAPPEDEDSAAFS	159
<input checked="" type="checkbox"/> AAA51780	76	SPRQ0000QGEDGSPQAHRRGPTGYLVLDEEQPSPQSALECHPERGCVPEPGAAVAASKGLPQQLPAPPEDEDSAAFS	155
<input checked="" type="checkbox"/> AAA51771	80	SPRQ0000QGEDGSPQAHRRGPTGYLVLDEEQPSPQSALECHPERGCVPEPGAAVAASKGLPQQLPAPPEDEDSAAFS	159

To evaluate sequences, change settings for “Conservation Setting” from “2 Bits” to “Identity”

**COBALT** Constraint-based Multiple Alignment Tool

Home Recent Results Help

Phylogenetic Tree Edit and Resubmit Download

- Cobalt RID EMV7SF1X211 (7 seqs)

All queries form only one cluster. No domain information was used for generating constraints. Decreasing maximum in-cluster distance or turning off query clustering option may improve results.

Descriptions Select All Re-align Alignment parameters

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Accession	Description	Links
<input checked="" type="checkbox"/> P10275.2	RecName: Full=Androgen receptor; AltName: Full=Dihydrotestosterone receptor; AltName: Full=Nucle	<a href="#">GM</a>
<input checked="" type="checkbox"/> AAA51772.1	androgen receptor [Homo sapiens] >gb AAA51771.1  androgen receptor [Homo sapiens]	<a href="#">G</a>
<input checked="" type="checkbox"/> AAA51780.1	androgen-receptor [Homo sapiens]	<a href="#">GM</a>
<input checked="" type="checkbox"/> AAA51771.1	androgen receptor [Homo sapiens] >gb AAA51772.1  androgen receptor [Homo sapiens]	<a href="#">GM</a>
<input checked="" type="checkbox"/> AAA51729.1	androgen receptor [Homo sapiens] >sp P10275.2 ANDR_HUMAN RecName: Full=Androgen receptor;	<a href="#">GM</a> <a href="#">PubChem BioAssay Info linked to AAA51729.1</a>
<input checked="" type="checkbox"/> AAD45921.1	androgen receptor [Homo sapiens]	<a href="#">GM</a>
<input checked="" type="checkbox"/> AAA51886.1	androgen receptor [Homo sapiens] >sp P10275.2 ANDR_HUMAN RecName: Full=Androgen receptor;	<a href="#">G</a> <a href="#">PubChem BioAssay Info linked to AAA51886.1</a>

Alignments Select All Re-align Mouse over the sequence identifier for sequence title

View Format: Compact Conservation Setting: Identity

<input checked="" type="checkbox"/> P10275	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000qET	80
<input checked="" type="checkbox"/> AAA51772	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000-ET	79
<input checked="" type="checkbox"/> AAA51780	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000----ET	75
<input checked="" type="checkbox"/> AAA51771	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000-ET	79
<input checked="" type="checkbox"/> AAA51729	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000qET	80
<input checked="" type="checkbox"/> AAD45921	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000----ET	75
<input checked="" type="checkbox"/> AAA51886	1	MEVQLGLGRVYFRPPSKTYRGAFQNLFQSVREVIQNPGRHPEAASAAPPGASLLLLLQ0000000000000000qET	80

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Look for differences in the sequence (e.g., conserved residues, gaps) and start by eliminating sequences that have gaps.

i. If, after the suggested evaluations of the proteins are performed, questions remain as to which sequence would be best to run in SeqAPASS, run all relevant sequences in SeqAPASS for the evaluation. The individual residue differences between commonly named sequences will become most important when evaluating residues known to be important for binding the chemical or activating the protein (Level 3 SeqAPASS analysis). After completing the SeqAPASS run, select the data that has the greatest number of ortholog candidates for your evaluation of conservation and further predictions of cross species susceptibility. Depending on the protein of interest, multiple subunits may be associated with a protein. In this case, all relevant subunits can be queried using SeqAPASS.

### *Level 1 Calculated Percent Similarity*

The SeqAPASS algorithms submit the query to NCBI's standalone BLASTp (using default settings, including BLOSUM-62 matrix), which aligns the query protein with all proteins available in the NCBI protein database and provides a variety of metrics associated with each pairwise alignment between the query and hit sequences. SeqAPASS selectively captures output from BLASTp, including one sequence per species with the highest bit score. Detailed descriptions of metrics derived from BLASTp (e.g., BLASTp Bitscore, E-Value, Positives, Identity, Hit length) can be found in:

The NCBI Handbook: (<http://www.ncbi.nlm.nih.gov/books/NBK21106/>);

BLAST® Help: (<http://www.ncbi.nlm.nih.gov/books/NBK62051/>) and the

NCBI Glossary Field Guide: (<http://www.ncbi.nlm.nih.gov/Class/FieldGuide/glossary.html>)

The top row of the Level 1 data corresponds to the queried protein selected by the user. For each sequence queried, the Level 1, top row query sequence is used to determine the maximum bitscore for the analysis, which is derived from aligning the query sequence to itself using BLASTp. To calculate percent similarity, the bitscore for each hit sequence is normalized to the maximum bit score and then multiplied by 100.

Note: SeqAPASS v2.0 and newer parse the BLASTp query and hit accessions to identify all the species/accessions from the identical proteins. Therefore, if a hit sequence represents multiple species, all species with the identical sequence will be found in the data table for Level 1 and Level 2. To determine which sequence/species was identified from BLASTp as a hit and which sequence/species was parsed from the identical sequence, view the "Full Report" for Level 1 or Level 2, column "Identical Protein," Where "N" is indicative of the original hit sequence and "Y" is the parsed sequence.

### *Common Domain Count*

Reversed Position Specific BLAST (RPS BLAST) is used to compare each query and hit sequence to conserved domains defined in NCBI's Conserved Domain Database. A hit domain is considered in common with the query domain if it contains the same domain accession as the query and it aligns with the NCBI curated domain with the same or greater amino acid residue coverage than the query sequence.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Ortholog Candidate Identification

Ortholog sequences are those that have diverged from a speciation event and therefore are more likely to maintain similar function. SeqAPASS uses reciprocal best hit (RBH) BLAST for ortholog detection by automatically comparing each hit protein to all protein sequences available for the query species and if the original query protein or one of its identical protein matches is identified to be the best match to the hit or maintain the same bitscore, then the hit sequence would be considered an ortholog candidate. The sequence is indicated an Ortholog Candidate or not with a yes (Y) or no (N) in the column.

**Note:** Many NCBI protein accessions represent multiple identical protein sequences in the BLASTp output. This is due to BLASTp querying and presenting data from the non-redundant protein database. Sometimes the identical sequences are from different species. This can be checked by following the link for the top row “NCBI Accession” in the table to the NCBI protein page. Below the protein name [species] title will be a link to “Identical Proteins.” Click the “Identical Proteins” link and look for a sequence in the list from the user defined query species.

The screenshot shows the NCBI protein page for 'estrogen receptor isoform 1 [Homo sapiens]'. At the top, there is a navigation bar with 'NCBI', 'Resources', and 'How To' links, along with a 'Sign in to NCBI' button. Below this is a search bar with a dropdown menu set to 'Protein' and a 'Search' button. A blue banner below the search bar contains a message: 'NCBI is phasing out sequence GI numbers in September 2016. Please use accession.version! Read more...'. The main content area shows the protein name 'estrogen receptor isoform 1 [Homo sapiens]' in bold, followed by the NCBI Reference Sequence: NP\_000116.2. There are links for 'Identical Proteins', 'FASTA', and 'Graphics'. On the right side, there are buttons for 'Change region shown' and 'Customize view', and a link for 'Analyze this sequence'.

**Note:** If the top hit is a Protein DataBank (PDB) code (e.g., 1AHR\_A) from RBH BLAST there will be no ortholog candidates identified. BLASTp when ran against all accessions for a given species does not return PDB codes. It is recommended that the user identify a similar/identical sequence to the PDB code and use that sequence as the query sequence.

### Susceptibility cut-off

The susceptibility cut-off values listed on the “Level 1 (and Level 2) Susceptibility Cut-off” page are determined by plotting the % similarity data from the “Primary Report” or “Full Report” and identifying the local minimums in the data. The default cut-off is determined by taking the 1<sup>st</sup> local minimum and moving up in percent similarity until the next ortholog candidate is found. The susceptibility cut-off displayed in the list is the percent similarity of the identified ortholog candidate.

*Criteria for Susceptibility Prediction (when “Primary Report Settings” is set to “Species Read-Across:” Yes)*

All sequences identified above the susceptibility cut-off are predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”

If the hit sequence is below the susceptibility cut-off, but identified as an Ortholog Candidate = Y, for “yes,” then the hit is predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

If the hit sequence is below the susceptibility cut-off but belongs to any organism class found above the susceptibility cut-off, the hit is predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”. This criterion allows susceptibility predictions to be made across taxonomic groups based on the likelihood that the sequences above the cut-off are better matches to the query.

If the hit sequence is below the susceptibility cut-off and not identified as an ortholog candidate (Ortholog Candidate = N, for “no,”) and does not belong to any organism class found above the susceptibility cut-off, the hit is predicted to not be susceptible; therefore, Susceptibility Prediction = N for “no”

Note that the “Primary Report” may yield different Susceptibility Predictions than the “Full Report,” as the predictions are based on the data in the different reports. The Primary Report is filtered to only display E-value  $\leq 0.01$  and Common Domain Count  $\geq 1$ .

*Criteria for Susceptibility Prediction (when “Primary Report Settings” is set to “Species Read-Across:” No)*

All sequences identified above the susceptibility cut-off are predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”

If the hit sequence is below the susceptibility cut-off, but identified as an Ortholog Candidate = Y, for “yes,” then the hit is predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”

If the hit sequence is below the susceptibility cut-off and not identified as an ortholog candidate (Ortholog Candidate = N, for “no,”), the hit is predicted to not be susceptible; therefore, Susceptibility Prediction = N for “no”

### *Level 2 Calculated Percent Similarity*

Data obtained from the Level 1 RPS BLAST evaluation is used to assign sequence ranges that aligned with a user selected domain (from the NCBI CDD database) to each accession from the Level 1 Full report. BLASTp is then used to align the query domain range to each hit domain range. The percent similarity is calculated based on the bit scores from the BLASTp alignment of the domain regions. For each sequence queried, the Level 2, top row query species is used to determine the maximum bitscore for the analysis, which is derived from aligning the query sequence to itself using BLASTp. To calculate percent similarity, the bitscore for each hit sequence is normalized to the maximum bit score and then multiplied by 100.

### *Susceptibility cut-off (same method as used in Level 1)*

The susceptibility cut-offs listed on the “Level 2 Susceptibility Cut-off” page are determined by plotting the % similarity data from the “Primary Report” or “Full Report” and identifying the local minimums in the data. The default cut-off is determined by taking the 1<sup>st</sup> local minimum and moving up in percent similarity until the next ortholog candidate is found. The susceptibility cut-off displayed in the list is the percent similarity of the identified ortholog candidate.

## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### *Level 2 Criteria for Susceptibility Prediction (when “Primary Report Settings” is set to “Species Read-Across:” Yes)*

All sequences identified above the susceptibility cut-off are predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”

If the hit sequence is below the susceptibility cut-off, but identified as an Ortholog Candidate = Y, for “yes,” then the hit is predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”

If the hit sequence is below the susceptibility cut-off but belongs to any organism class found above the susceptibility cut-off, the hit is predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”. This criterion allows susceptibility predictions to be made across taxonomic groups based on the likelihood that the sequences above the cut-off are better matches to the query.

If the hit sequence is below the susceptibility cut-off and not identified as an ortholog candidate (Ortholog Candidate = N, for “no,”) and does not belong to any organism class found above the susceptibility cut-off, the hit is predicted to not be susceptible; therefore, Susceptibility Prediction = N for “no”

Note that the “Primary Report” may yield different Susceptibility Predictions than the “Full Report,” as the predictions are based on the data in the different reports. The Primary Report is filtered to only display E-value  $\leq 0.01$  and Common Domain Count  $\geq 1$ .

### *Level 2 Criteria for Susceptibility Prediction (when “Primary Report Settings” is set to “Species Read-Across:” No)*

All sequences identified above the susceptibility cut-off are predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”

If the hit sequence is below the susceptibility cut-off, but identified as an Ortholog Candidate = Y, for “yes,” then the hit is predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”

If the hit sequence is below the susceptibility cut-off and not identified as an ortholog candidate (Ortholog Candidate = N, for “no,”), the hit is predicted to not be susceptible; therefore, Susceptibility Prediction = N for “no”

### *Level 3 Sequence Alignments*

COBALT is used to align all user selected sequences (from Level 1 hits) with a user defined template sequence. Because COBALT algorithms align all sequences, it is recommended that the user align the template sequence with sequences that are most similar to one another. As a means to capture the most similar sequences from the SeqAPASS data it is recommended that the user filter the Level 1 data by taxonomic group and step through the Level 1 data pages one by one while selecting sequences. It is recommended that the user look at the name of the sequence and exclude ‘partial’ sequences when possible. Requesting a query from one taxonomic group at a time, breaks the data down in manageable alignments.

### *Selecting Amino Acid Residues to Align*

The user may select up to 50 amino acid residues to compare across selected species in Level 3.

## **Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide**

Updated 08/14/2023; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### *Level 4 Structural Alignments*

Level 4 is intended for the use by expert users only. It is recommended that each user only submit ten I-TASSER jobs at a time and wait for them to complete prior to submitting more. Although it is inconvenient the user must begin the Level 4 request process over beginning with submitting a user defined name for the job, generating FASTA, and submitting the next 10 FASTA to I-TASSER. A recommended naming format could be by Protein accession followed by 1\_10; Protein accession 11-20, etc. The user should prioritize which jobs they need completed first. The I-TASSER jobs take days to complete, therefore it is important for the expert users of the tool to not overwhelm the system by submitting too many jobs at once. Proteins that have been submitted to I-TASSER previously within a query will be marked as submitted. The user should pay attention to which I-TASSER jobs have been submitted previously.

### *TM-align*

The TM-align susceptibility prediction is “Y”, indicating yes, by default. It is intended for the expert user to determine whether that prediction is a yes or a no.